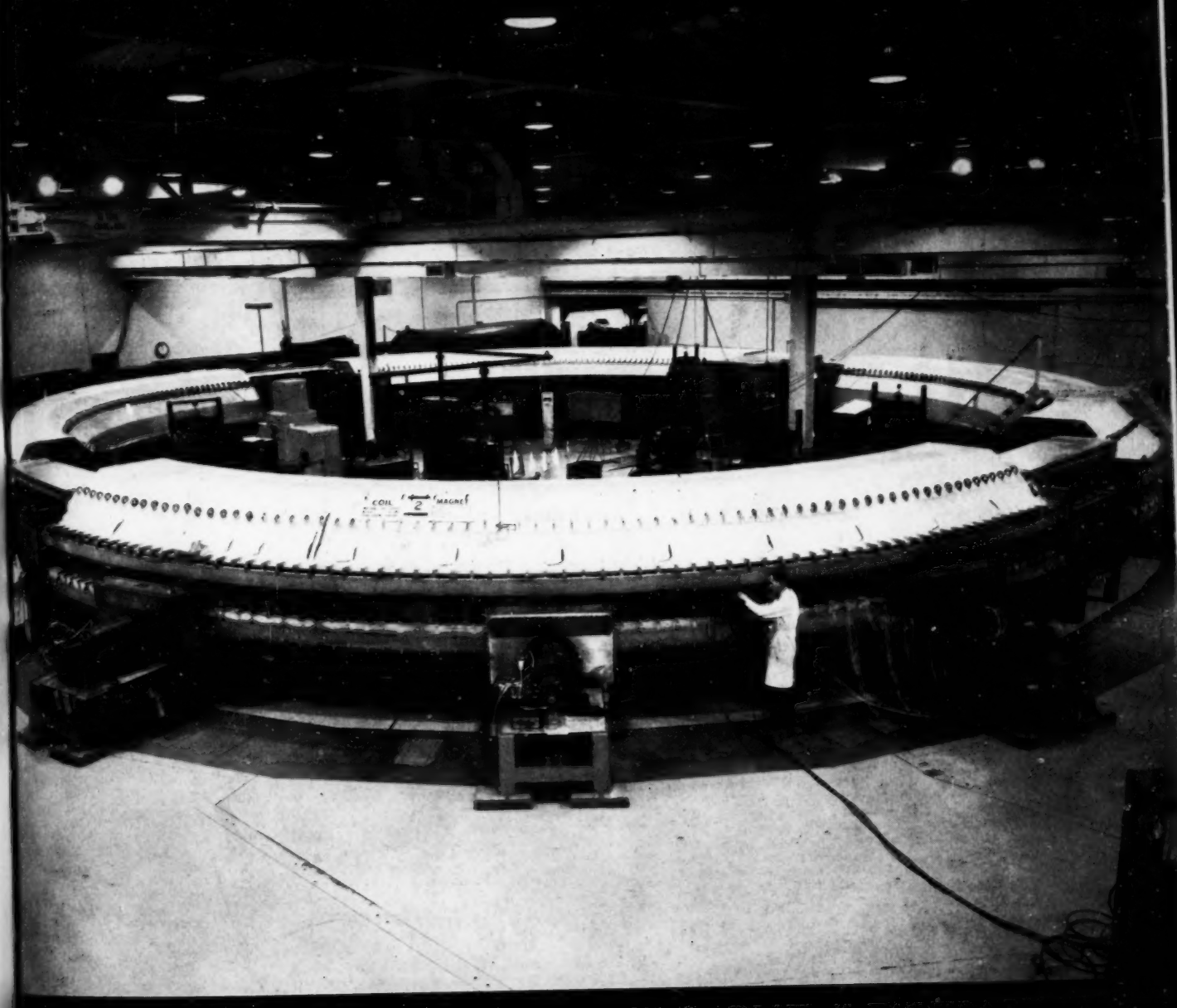
JOURNAL OF APPLIED PHYSICS



JOURNAL
OF
APPLIED PHYSICS

Vol. 23, No. 7 JULY, 1952

In This Issue

		Page
Completeness Relations for Loss-Free Microwave Junctions	T. Teichmann	701
Methods of Measuring the Properties of Ionized Gases at High Frequencies. I. Measurements of *Q*	Sanborn C. Brown and David J. Rose	711
Methods of Measuring the Properties of Ionized Gases at High Frequencies. II. Measurement of Electric Field	David J. Rose and Sanborn C. Brown	719
The Initial Conduction Interval in High Speed Thyratrons	James B. Woodford, Jr., and Everard M. Williams	722
On Diffusionless Transformation in Au-Cd Single Crystals Containing 47.5 Atomic Percent Cadmium: Characteristics of Single-Interface Transformation	Lo-Ching Chang	725
Specimen Charging in the Electron Microscope and Some Observations on the Size of Polystyrene Latex Particles	S. G. Ellis	728
Theory of Wave-Guide-Fed Slots Radiating into Parallel-Plate Regions	Harry Gruenberg	733
On the Evaluation of Noise Samples	Arnold J. F. Siegert	737
Admittance of a Diode with a Retarding Field	J. J. Freeman	743
Creep of Zinc Single Crystals	L. Slifkin and W. Kauzmann	746
Theory of an Oscillating Cylinder Viscometer. Part I	Ali A. K. Ibrahim and A. M. Kabiel	754
Measurements of the Physical Properties of Active Nitrogen	James M. Benson	757
Experimental Verification of Source Size Theory for the Mach-Zehnder Interferometer	G. D. Kahl and F. D. Bennett	763
Dielectric Properties of Ice and Snow at 3.2 Centimeters	W. A. Cumming	768
Axial Motion of an Electron in a Constant Wave Velocity Section of a Linear Accelerator	David Caplan and Edward Akeley	774
Electrical Resistance of Binary Metallic Mixtures	Rolf Landauer	779
Grain Boundary Diffusion of Zinc in Copper	R. Flanagan and R. Smoluchowski	785
Use of Radioactive Isotopes in a Study of Evaporation from Thermionic Cathodes	W. F. Leverton and W. G. Shepherd	787
Reliability of Internal Standards for Calibrating Electron Miscroscopes	John H. L. Watson and William L. Grube	793
Letters to the Editors:		
Note on the Strain Aging of Iron Single Crystals	A. N. Holden and F. W. Kunz	799
A High Vacuum Flow Analogy	G. W. Monk	799
Crack Formation on Sodium Chloride Single Crystal Surfaces	Robert A. Lad	800
Note on the Discharge Coefficient of a Critical-Flow Nozzle	Henry Wise	801
On the Redundant Information Supplied in Practical Applications by the Time and Frequency-Phase Responses of a System	M. Levy	801
Random Noise in Dielectrics	Herbert Bauss and W. F. Boyer	802
Announcements		803
Index to Advertisers		xxiv

BOARD OF EDITORS
Elmer Hutchisson,
Editor
Earle C. Gregg,
Assistant Editor

Associate Editors
Howard H. Aiken
Ray E. Bolz
Lyle D. Borst
Robert D. Heidenreich
James Hillier
Leonard B. Loeb
W. James Lyons
Eric Reissner
Roman Smoluchowski

Publication of the
American Institute of Physics

OFFICERS
George R. Harrison,
Chairman
George B. Pegram,
Treasurer

GOVERNING BOARD
Karl K. Darrow
G. J. Dienes
S. A. Goudsmit
G. P. Harnwell
George R. Harrison
F. V. Hunt
Rudolph Kingslake
Hugh S. Knowles
C. C. Lauritsen
F. W. Loomis
William F. Meggers
George B. Pegram
I. I. Rabi
Duane Roller
Frederick Seitz
Francis G. Slack
John C. Steinberg
Richard M. Sutton
Mark W. Zemansky

Consultant on Printing
Melvin Loos

ADMINISTRATIVE STAFF
Henry A. Barton,
Director
Wallace Waterfall,
Executive Secretary
Theodore Vorburger,
Advertising Manager
Ruth F. Bryans,
Publication Manager
Betty H. Goodfriend,
Assistant Secretary
Eileen B. Neuberger,
Circulation Manager
Kathryn Setze
Assistant Treasurer

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials and reviews, as well as technical papers of applied physics. Articles appearing in it are indexed regularly, according to their field, by Physics Abstracts, Chemical Abstracts, Engineering Index, Inc., Industrial Arts Index, etc.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio. Unless otherwise stated, submission of a manuscript is a representation that it has been neither copyrighted, published, nor submitted for publication elsewhere. The authors' institutions are requested to pay a **publication charge** of $8 per page which, if honored, entitles them to 100 reprints without further charge. Instructions will be sent with galley proofs.

Subscriptions, renewals, and orders for back numbers should be addressed to the American Institute of Physics, 57 East 55 Street, New York 22, New York.

Subscription Price	U.S. and Canada	Elsewhere
To members of the American Institute of Physics and Affiliated Societies	$10.00	$11.00
To all others	12.00	13.00

Back Number Prices

Yearly back number rate when complete year is available: $14.00. Single copies: $1.25 each.

Changes of address should be sent to the American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics. **Advertising rates** supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in paragraph (d-2), section 34.40, P. L. & R. of 1948, authorized May 2, 1932.

Journal
of
Applied Physics

Volume 23, Number 7 July, 1952

Completeness Relations for Loss-Free Microwave Junctions

T. TEICHMANN
Palmer Physical Laboratory, Princeton University, Princeton, New Jersey
(Received February 25, 1952)

Using a method which is the electromagnetic analog of the scattering matrix formalism in the theory of nuclear reactions, two "sum rules" are derived for the frequency independent coefficients occurring in the admittance matrix relating the currents and voltages in a loss-free microwave junction. These are sums, respectively, over the various modes of the guides entering the junction, and over the modes of the cavity comprising the junction (suitably defined). The results are used to estimate the effect on the admittance matrix of the higher modes, both of the guides, and of the junction proper.

I. INTRODUCTION

THE formal similarity between the theory of nuclear reactions and that of "scattering" of electromagnetic waves in microwave junctions seems to have been noticed together with the development of both these subjects during recent years. In particular, mention is made of this point by Wigner,[1] and some papers by Tomonaga[2,3] and collaborators[4] develop basic aspects of the theory of microwave junctions using methods more characteristic of nuclear physics.[5]

There is no published work extant, however, which uses results and methods of one of these disciplines to derive analogous and unfamiliar results of the other. This is readily understandable, for in the electromagnetic case it has generally been possible, at least in principle, to calculate to the desired order of approximation the fields, waves, and what not, for a given system of guides and junction, at any rate for the lowest modes. The most general formal features are known, but otherwise each case is treated on its own merits. In the nuclear case, on the other hand, detailed calculations are quite impossible at the present time, except on the basis of very unrealistic models, and it has been necessary to drive further the formal approach, using only the most general features of the boundary value problem involved.[6,7,8]

It is the aim of this paper to derive *completeness relations* for microwave junctions, corresponding to certain *sum rules* for nuclear reactions. Since the nuclear method does not single out the lowest states for special treatment, it is clear from the remarks above that the analogous results in the electromagnetic case can be expected to yield new information mainly about the higher modes, both of the junction and the guides.

The treatment here is confined to the loss-free case; similar, though more complicated results can be derived for the lossy case, but they will not be dealt with here.

II. THE ELECTROMAGNETIC FIELD INSIDE THE GUIDES AND JUNCTION: THE ADMITTANCE AND COLLISION MATRICES

The results to be derived below concern sums of certain constant (i.e., frequency independent) coefficients occurring in the admittance matrix Y, and in order to clarify the notation and the nature of the quantities involved, a sketch of the determination of Y is given below, together with some relevant subsidiary remarks. More detailed derivations, and a full discussion may be found in articles by Fränz[9] and Slater.[10]

[1] E. P. Wigner, Am. J. Phys. **17**, 99 (1949).
[2] S. Tomonaga, J. Phys. Soc. (Japan) **2**, 158 (1947).
[3] S. Tomonaga, J. Phys. Soc. (Japan) **3**, 93 (1948).
[4] T. Miyazima and T. Tati, J. Sci. Research Inst. (Tokyo) **43**, 1 (1949). I am grateful to Professor Tomonaga for sending me this, and several other papers which were not available here.
[5] From the notational point of view the similarity is also brought out in some unpublished wartime reports of H. A. Bethe. I am indebted to Dr. H. Motz for bringing these to my attention.
[6] E. P. Wigner and L. Eisenbud, Phys. Rev. **72**, 29 (1947).
[7] T. Teichmann, Ph.D. thesis, Princeton University (1949).
[8] T. Teichmann and E. P. Wigner, (to be published).
[9] K. Fränz, Elek. Nachr.-Tech. **21**, 8 (1944).
[10] J. C. Slater, Revs. Modern Phys. **18**, 441 (1946). This reference is particularly adapted to the needs of the present paper. For my own background I am, however, mainly indebted to some unpublished manuscripts of Professor R. H. Dicke and Professor E. P. Wigner.

One considers a set of wave guides with perfectly conducting walls, joined by a cavity, hereafter called the junction, which also has perfectly conducting walls. The material inside the guides and junction is assumed to be perfectly insulating, but may have arbitrary (linear) electric and magnetic characteristics. For simplicity ϵ and μ will be taken as constant, but a study of the following arguments shows that the extension to the case where ϵ and μ are arbitrary functions of position is not very difficult.

It will also be assumed that the guides are hollow, and that the junction is a simply connected body when the guide openings are closed. This assumption is made to avoid various mathematical and physical complications, which do not affect the general nature of the results here aimed at in any essential way, despite their intrinsic and in some cases even practical interest. This assumption of simple connectivity excludes TEM modes in the guides, and quasi-static electric modes in the junction. The presence of the guide openings, however, makes it necessary to consider quasi-static magnetic fields in the junction: These play a role somewhat similar to that of the near field of an antenna, and which can be minimized by a suitable choice of the terminal planes of the guides (see below).[11]

For reasons which are apparent, one introduces a plane T_N in each wave guide "N," perpendicular to its generators (defined by the unit vector $\mathbf{e}_N$), which is called the *terminal plane.* (See Fig. 1.) These planes serve to delimit the junction and the guides for mathematical purposes: that part of each guide to the junction side of the terminal plane is regarded as part of the junction. The junction thus consists of a volume V limited by a boundary $S+T$, S being the physical (perfectly conducting) part of the boundary, and $T=T_1+T_2+T_3+\cdots$ being the artificial part, comprising the portions T_1, T_2, T_3, $\cdots$ of the terminal planes intercepted by the perfectly conducting walls $C_1, C_2, C_3, \cdots$ of the guides $1, 2, 3, \cdots$. The symbol $T_N, N=1, 2, 3, \cdots$ will in the remainder of this article denote only that part of the terminal plane *within* the guide N, and is thus simply a guide cross section. Its boundary (i.e., its intersection with C_N) will be denoted by c_N. It is clear that the introduction of these terminal planes constitutes an arbitrary element in the theory, though for any reasonably shaped cavity physical considerations may be expected to dictate a natural position. Since these planes will have to be moved from this position if one wishes to minimize, or at least simplify, the effect of the quasi-static field in the junction, it is satisfying to know that it is always possible to calculate, to any order of approximation, the effect on the parameters, (introduced below), of any variation in the position of the planes. [See, for example, Slater's discussion of the perturbation of boundary conditions (reference 10), Tomonaga's description of circuits with pistons (reference 2), and the author's treatment of the analogous problem in the theory of nuclear reactions.]

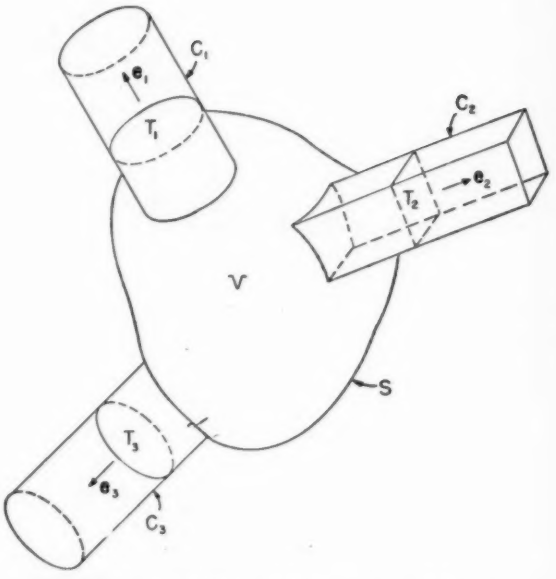


FIG. 1. Schematic diagram of a wave-guide junction.

If the electromagnetic field is subjected to a temporal Fourier analysis, the components of frequency ω satisfy the equations

$$\begin{aligned} \text{curl}\mathbf{E} &= -j\omega\mu\mathbf{H} \\ \text{curl}\mathbf{H} &= j\omega\mathbf{E}\epsilon \end{aligned} \tag{2.1}$$

$$\begin{aligned} \text{div}\epsilon\mathbf{E} &= 0 \\ \text{div}\mu\mathbf{H} &= 0 \end{aligned} \tag{2.2}$$

in the medium, and the boundary conditions

$$\mathbf{E}\times\mathbf{n}=0 \quad \mathbf{H}\cdot\mathbf{n}=0 \tag{2.3}$$

on the perfectly conducting surface $S+C_1+C_2+C_3+\cdots=S+C$, $\mathbf{n}$ being the normal to this surface.

Because of the cylindrical nature of the wave guides, the field (of frequency ω) in the Nth guide, may be written in the form

$$\begin{aligned} \mathbf{E}_N &= \sum_s a_{N_s}\mathbf{F}_{N_s}^{-}\exp(-jk_{N_s}z_N) \\ &\qquad + b_{N_s}\mathbf{F}_{N_s}^{+}\exp(jk_{N_s}z_N) \\ H_N &= \sum_s a_{N_s}\mathbf{G}_{N_s}^{-}\exp(-jk_{N_s}z_N) \\ &\qquad + b_{N_s}\mathbf{G}_{N_s}^{+}\exp(jk_{N_s}z_N), \end{aligned} \tag{2.4}$$

where $\mathbf{F}_{N_s}^{+-}$, and $\mathbf{G}_{N_s}^{+-}$ are vectors which are functions only of the cross-sectional coordinates of the guide (i.e., of the position in the terminal plane T_N). z_N denotes the distance along the guide in the direction $\mathbf{e}_N$, measured from the terminal plane. a_{N_s} and b_{N_s} are frequency dependent constants, and depending on their definition, $\mathbf{F}_{N_s}^{+-}$ and $\mathbf{G}_{N_s}^{+-}$ also depend on ω. k_{N_s}, which will be defined below, also depends on frequency, but not on position.

[11] E. P. Wigner (unpublished).

It is now most convenient to distinguish transverse vectors $\mathbf{A}_{t,N}$, which have no component along the guide, and thus satisfy

$$\mathbf{A}_{t,N}\cdot\mathbf{e}_N=0 \tag{2.5}$$

and longitudinal vectors $\mathbf{A}_{z,N}$, which have no components parallel to the terminal plane, and thus satisfy

$$\mathbf{A}_{z,N}\times\mathbf{e}_N=0. \tag{2.6}$$

Any vector may then be uniquely resolved into the sum of a transverse and a longitudinal vector

$$\mathbf{A}_N=\mathbf{A}_{t,N}+\mathbf{A}_{z,N} \tag{2.7}$$

with

$$\mathbf{A}_{t,N}\cdot\mathbf{A}_{z,N}=0. \tag{2.8}$$

Introducing this resolution for $\mathbf{F}_{Ns}$ and $\mathbf{G}_{Ns}$, the field in the guide may be rewritten

$$\begin{aligned}\mathbf{E}_N&=\sum_s v_{Ns}(z_N)\mathbf{F}_{t,Ns}+i_{Ns}(z_N)\mathbf{F}_{z,Ns}\\ \mathbf{H}_N&=\sum_s i_{Ns}(z_N)\mathbf{G}_{t,Ns}+v_{Ns}(z_N)\mathbf{G}_{z,Ns},\end{aligned} \tag{2.9}$$

where $v_{Ns}(z_N)$ and $i_{Ns}(z_N)$ represent the generalized voltages and currents, and satisfy the relations

$$\begin{aligned}v_{Ns}(z_N)&=a_{Ns}\exp(-jk_{Ns}z_N)+b_{Ns}\exp(jk_{Ns}z_N)\\ i_{Ns}(z_N)&=a_{Ns}\exp(-jk_{Ns}z_N)-b_{Ns}\exp(jk_{Ns}z_N)\end{aligned} \tag{2.10}$$

(see reference 10). $\mathbf{F}_{Ns}$ and $\mathbf{G}_{Ns}$ are defined in terms of the solutions U_{Ns} of the following scalar boundary value problem

$$\operatorname{grad}_{t,N}{}^2U_{Ns}+K_{Ns}{}^2U_{Ns}=0 \quad \text{on} \quad T_N \tag{2.11}$$

with corresponding proper values

$$K_{Ns}{}^2=\omega^2\mu\epsilon-k_{Ns}{}^2 \tag{2.12}$$

which define k_{Ns}, with the boundary condition

$$\begin{aligned}U_{Ns}&=0 \quad \text{on } c_N \text{ for even } s\\ (\operatorname{grad}_{t,N}U_{Ns})_n&=0 \quad \text{on } c_N \text{ for odd } s.\end{aligned} \tag{2.13}$$

These U_{Ns} thus form two complete sets of orthogonal functions, one each for even s and odd s, and have been enumerated as one set merely for convenience. They satisfy the orthogonality relations

$$\begin{aligned}\int_{T_N}U_{Nr}U_{Ns}dT_N&=\int_{T_N}(\operatorname{grad}_{t,N}U_{Nr})\\ &\qquad\cdot(\operatorname{grad}_{t,N}U_{Ns})dT_N\\ &=0 \quad \text{for} \quad r=s-2m, \quad m\neq 0\end{aligned} \tag{2.14}$$

i.e., for $r\neq s$ and r, s either both even or both odd. They will here be normalized so that

$$\begin{aligned}\int_{T_N}(\operatorname{grad}_{t,N}U_{Ns})^2dT_N&=K_{Ns}{}^2\int_{T_N}U_{Ns}{}^2dT_N\\ &=1.\end{aligned} \tag{2.15}$$

There is no loss of generality in choosing them real.

The two types of modes are then given as follows:

TM, i.e., *E*-modes, s even:

$$\begin{aligned}\mathbf{F}_{t,Ns}&=(2\omega Z_{Ns})^{\frac{1}{2}}\operatorname{grad}_{t,N}U_{Ns}\\ \mathbf{F}_{z,Ns}&=(2\omega Z_{Ns})^{\frac{1}{2}}(jK_{Ns}{}^2/k_{Ns})U_{Ns}\mathbf{e}_N\\ \mathbf{G}_{t,Ns}&=(\mathbf{F}_{t,Ns}\times\mathbf{e}_N)/Z_{Ns}\\ \mathbf{G}_{z,Ns}&=0\end{aligned} \tag{2.16}$$

TE, i.e., *H*-modes, s odd:

$$\begin{aligned}\mathbf{G}_{t,Ns}&=(2\omega/Z_{Ns})^{\frac{1}{2}}\operatorname{grad}_{t,N}U_{Ns}\\ \mathbf{G}_{z,Ns}&=(2\omega/Z_{Ns})^{\frac{1}{2}}(K_{Ns}{}^2/jk_{Ns})U_{Ns}\mathbf{e}_N\\ \mathbf{F}_{t,Ns}&=Z_{Ns}(\mathbf{e}_N\times\mathbf{G}_{t,Ns})\\ \mathbf{F}_{z,Ns}&=0.\end{aligned} \tag{2.17}$$

Z_{Ns} is the impedance of the mode considered, and may be most compactly described by the equation

$$Z_{Ns}=\left(\frac{\mu}{\epsilon}\right)^{\frac{1}{2}}\left(1-\frac{K_{Ns}{}^2}{\omega^2\mu\epsilon}\right)^{\frac{1}{2}(-1)^s}. \tag{2.18}$$

It should be noted that at any fixed frequency ω only a certain number of modes are "allowed," i.e., maintain their amplitude, and transfer any appreciable energy along the guide. This follows from Eq. (2.12): $K_{Ns}{}^2$ is a monotonic nondecreasing function of s (in the usual sort of enumeration), and hence k_{Ns} will be real only for $s\leqslant s_0=s_0(\omega)$, s_0 being the lowest s for which

$$K_{Ns}{}^2<\omega^2\mu\epsilon \tag{2.19}$$

(s_0 varies from guide to guide, i.e., depends on N). For $s\leqslant s_0$ the energy flux along the guide for a wave of unit amplitude travelling inward is

$$\begin{aligned}P_{Ns}&=(1/2)\int_{T_N}(\mathbf{E}_{Ns}\times\mathbf{H}_{Ns})\cdot\mathbf{e}_N dT_N\\ &=\omega.\end{aligned} \tag{2.20}$$

For $s>s_0$, on the other hand, k_{Ns} is imaginary, and the corresponding fields decay exponentially down the guide. The corresponding "flux" then also decays exponentially, and has the value $j\omega$ at the initial point, showing that no power is transferred down the guide. Equation (2.20) shows that the normalization has been so chosen as to give unit flux per (angular) cycle for the allowed modes.

It seems worthwhile to remark, parenthetically, that these and the following considerations may be extended to guides which are not cylindrical, provided that Maxwell's equations are separable in such a way that the field may be represented as a sum of products of "cross-sectional" vectors, corresponding to $\mathbf{F}_{Ns}$, $\mathbf{G}_{Ns}$ above, and incoming or emitted waves, corresponding to $\exp(\pm jk_{Ns}z_N)$, but not necessarily plane. It is probably unnecessary to point out that the terminal planes will then have to be replaced by terminal surfaces of a shape appropriate to the separated coordinates.

The complete field outside the junction may thus be written

$$\mathbf{E}=\sum_N \mathbf{E}_N=\sum_N \sum_s v_{Ns}(z_N)\mathbf{F}_{t,Ns}+i_{Ns}(z_N)\mathbf{F}_{z,Ns}$$
$$\mathbf{H}=\sum_N \mathbf{H}_N=\sum_N \sum_s i_{Ns}(z_N)\mathbf{G}_{t,Ns}+v_{Ns}(z_N)\mathbf{G}_{z,Ns}. \tag{2.21}$$

The perfect conductivity ensures that the vectors $\mathbf{F}_{Ns}$, $\mathbf{G}_{Ns}$ are zero everywhere except in the Nth guide. This holds to a very high degree of approximation even for ordinary conductors, unless the walls of the guides are vanishingly thin. The condition to be satisfied is that, at the frequency considered, the walls of the guide should be thicker than the corresponding skin depth for the material of the wall. As a result one has orthogonality relations not only between different modes in the same guide, but also between any two modes in different guides. The relevant orthogonality relations are thus

$$\int_{T_N} \mathbf{F}_{t,Lr}\cdot\mathbf{G}_{t,Ms}dT_N$$
$$=\int_N (\mathbf{F}_{t,Lr}\times\mathbf{G}_{t,Ms})\cdot\mathbf{e}_N dT_N$$
$$=2\omega\delta_{NL}\delta_{NM}\delta_{rs}. \tag{2.22}$$

All other integrals of this type can be obtained simply from the above by reference to Eqs. (2.15) to (2.17).

In order to describe the field inside the junction, it is necessary to introduce a complete, and preferably orthogonal, set of functions (vectors) inside the region V. In the case of a completely closed cavity, such a set may be obtained by a method described in a famous paper by Weyl,[12] in which the effects of multiple connectivity were also considered. The functions thus obtained satisfy homogeneous "short circuit" boundary conditions on the entire surface $S+T$ of the junction (i.e., $\mathbf{E}\times\mathbf{n}=0$ on $S+T$), and while they are basic to the expansion of the fields of interest here, they are *not* complete with respect to these fields. It is necessary to add an irrotational (i.e., longitudinal) magnetic field for completeness.[13] For some purposes one might wish to use fields satisfying homogeneous "open circuit" boundary conditions (i.e., $\mathbf{H}\times\mathbf{n}=0$), and in this case it is necessary to add an irrotational electric field instead of a magnetic field. Formally it is possible to circumvent these irrotational fields by imposing an *impedance* boundary condition at the terminal planes T

$$\mathbf{E}\times\mathbf{n}=\mathfrak{A}\cdot(\mathbf{H}\times\mathbf{n}), \tag{2.23}$$

where $\mathfrak{A}$ is a two-dimensional dyadic with zero trace, while retaining the short circuit conditions on the conducting surface S. This leads to some formal simplification in the derivation of the form of the admittance matrix Y; Y itself is, however, rather complicated, and not well adapted to practical calculations. In the case of homogeneous boundary conditions, the nonirrotational part of Y has a much simpler form, and the contribution of the irrotational part too can be reduced to a very simple form by suitable choice of the terminal planes. Since it is also much easier to calculate the normal modes of a cavity with homogeneous boundary conditions, such conditions will be imposed here, despite the slight additional complication introduced in the initial part of the derivation by the presence of an irrotational magnetic field. It should be noted that this irrotational field arises in the limit $\mathfrak{A}\equiv 0$, because an infinity of normal modes drop out (coalesce with others), and the remainder, though still infinite, are no longer complete with respect to the field to be expanded.

The non-irrotational part of the field inside the junction may thus be characterized by a set of proper fields $\mathbf{E}_\alpha$, $\mathbf{H}_\alpha$ of frequency ω_α [i.e., time dependence $\exp(j\omega_\alpha t)$], satisfying the equations

$$\begin{aligned}\operatorname{curl}\mathbf{E}_\alpha &= -j\omega_\alpha\mu\mathbf{H}_\alpha\\ \operatorname{curl}\mathbf{H}_\alpha &= j\omega_\alpha\epsilon\mathbf{E}_\alpha\end{aligned} \tag{2.24}$$

and the boundary condition

$$\mathbf{E}_\alpha\times\mathbf{n}=0 \quad \text{on} \quad S+T. \tag{2.25}$$

Since

$$\operatorname{div}(\mathbf{E}_\alpha\times\mathbf{H}_\beta-\mathbf{E}_\beta\times\mathbf{H}_\alpha)$$
$$=j(\omega_\alpha-\omega_\beta)(\epsilon\mathbf{E}_\alpha\cdot\mathbf{E}_\beta-\mu\mathbf{H}_\alpha\cdot\mathbf{H}_\beta), \tag{2.26}$$

one finds from Gauss' theorem that

$$j(\omega_\alpha-\omega_\beta)\int_V(\epsilon\mathbf{E}_\alpha\cdot\mathbf{E}_\beta-\mu\mathbf{H}_\alpha\cdot\mathbf{H}_\beta)dV$$
$$=\int_{S+T}(\mathbf{E}_\alpha\times\mathbf{H}_\beta-\mathbf{E}_\beta\times\mathbf{H}_\alpha)\cdot\mathbf{n}dS$$
$$=0 \tag{2.27}$$

by virtue of the boundary condition (2.25). Applying similar considerations to Eqs. (2.24) and their complex conjugates, one finds directly that

(1) the ω_α are real;

(2) $\mathbf{E}_\alpha$ may be chosen real, $\mathbf{H}_\alpha$ imaginary; (2.28)

(3) solutions occur in pairs, $\mathbf{E}_\alpha$, $\mathbf{H}_\alpha$, ω_α and $\mathbf{E}_{-\alpha}=\mathbf{E}_\alpha$, $\mathbf{H}_{-\alpha}=-\mathbf{H}_\alpha$, $\omega_{-\alpha}=-\omega_\alpha$.

It is, therefore, possible to normalize the $\mathbf{E}_\alpha$, $\mathbf{H}_\alpha$ in such a way that

$$\int_V(\epsilon\mathbf{E}_\alpha\cdot\mathbf{E}_\beta-\mu\mathbf{H}_\alpha\cdot\mathbf{H}_\beta)dV$$
$$=\int_V(\epsilon\mathbf{E}_\alpha\cdot\mathbf{E}_\beta^*+\mu\mathbf{H}_\alpha\cdot\mathbf{H}_\beta^*)dV$$
$$=\delta_{\alpha\beta}, \tag{2.29}$$

[12] H. Weyl, J. f. Math. **143**, 177 (1913). The main object of this paper was, however, the derivation of the asymptotic distribution of the proper frequencies of the cavity.

[13] Details are given in a forthcoming note by E. P. Wigner and the author; the lack of completeness seems to have been pointed out first by E. P. Wigner in reference 11.

where the star denotes complex conjugation. Since $\mathbf{E}_\alpha$, $\mathbf{H}_\alpha$ each satisfy the vector Helmholtz equation, it is clear that one also has

$$2\int_V \epsilon \mathbf{E}_\alpha \cdot \mathbf{E}_\beta dV = -2\int_V \mu \mathbf{H}_\alpha \cdot \mathbf{H}_\beta dV = \delta_{\alpha\beta}. \tag{2.30}$$

The field $\mathbf{E}$, $\mathbf{H}$ satisfying Eqs. (2.1), (2.2), and (2.3) in the junction and on the conducting surface S, may then be expanded in the series

$$E=\sum_\alpha f_\alpha \mathbf{E}_\alpha, \qquad H=\sum_\alpha f_\alpha \mathbf{H}_\alpha + \text{grad}u, \tag{2.31}$$

where u satisfies Laplace's equation

$$\text{grad}^2 u = 0 \tag{2.32}$$

and the boundary condition

$$\mathbf{n}\cdot\text{grad}u = 0 \quad \text{on } S. \tag{2.33}$$

Such a vector grad$u$ is orthogonal to all the $\mathbf{H}_\alpha$, but not to an arbitrary $\mathbf{H}$ excited by the wave-guide openings. Using the analog of Eq. (2.26), Gauss' theorem, and the boundary condition (2.25), one finds that

$$j(\omega_\alpha-\omega)f_\alpha = \int_T \mathbf{H}_\alpha \cdot (\mathbf{E}_t \times \mathbf{n}) dT. \tag{2.34}$$

On T, $\mathbf{E}$ and $\mathbf{H}$ may be expanded in terms of the wave guide modes [Eq. (2.21)] so that $\mathbf{E}_t = \sum_N \sum_s v_{Ns}\mathbf{F}_{t,Ns}$ and $\mathbf{H}_t = \sum_N \sum_s i_{Ns}\mathbf{G}_{t,Ns}$, v_{Ns} and i_{Ns} being evaluated for $z_N = 0$. Thus

$$(\omega_\alpha-\omega)f_\alpha = \sum_N \sum_s v_{Ns}(2\omega Z_{Ns})^{\frac{1}{2}}\gamma_{\alpha Ns}, \tag{2.35}$$

where the quantity

$$\gamma_{\alpha Ns} = -j(Z_{Ns}/2\omega)^{\frac{1}{2}} \int_{TN} \mathbf{H}_\alpha \cdot \mathbf{G}_{t,Ns} dT_N \tag{2.36}$$

is real and independent of frequency. On the other hand, the orthogonality relations (2.22) show that

$$\int_{T_M} \mathbf{H}\cdot\mathbf{G}_{t,Mr} dT_M = (2\omega/Z_{Mr})i_{Mr} = \sum_\alpha j f_\alpha (2\omega/Z_{Mr})^{\frac{1}{2}}\gamma_{\alpha Mr} + \int_{T_M} \mathbf{G}_{t,Mr}\cdot\text{grad}u\, dT_M. \tag{2.37}$$

Introducing the matrix

$$A_{MrNs} = j\sum_\alpha \frac{\gamma_{\alpha Mr}\gamma_{\alpha Ns}}{\omega_\alpha - \omega} \tag{2.38}$$

the sum in (2.37) becomes

$$2\omega \sum_N \sum_s Z_{Mr}^{-\frac{1}{2}} A_{MrNs} Z_{Ns}^{\frac{1}{2}} v_{Ns}.$$

In order to evaluate the integral, one uses the fact that u may be expressed in terms of its normal derivative over T by means of the Neumann function, $N(p, q)$, by the formula[14]

$$u(p) = \int_T N(p, q)\mathbf{n}\cdot\text{grad}u(q) dT(q). \tag{2.39}$$

On T the normal component of the field H is given by $\sum_N \sum_s v_{Ns}G_{z,Ns}$, and hence the integral in Eq. (2.37) becomes

$$\sum_N \sum_s -2j\omega v_{Ns}(Z_{Mr}Z_{Ns})^{-\frac{1}{2}} k_{Ns}^{-1} K_{Mr}^2 K_{Ns}^2 \times \int_{T_M}\int_{T_N} dT_M(p) dT_N(q) U_{Mr}(p) N(p, q) U_{Ns}(q).$$

This result follows quite directly from Eqs. (2.37) and (2.39), by applying Green's theorem to the integral, and then using formulas (2.11), (2.13), and (2.17). The double integral can be evaluated to give zero if r or s or both are electric modes, while if they are *both* magnetic modes it is given by $K_{Mr}^{-3}\delta_{MN}\delta_{rs}$ plus another term which is less than $B_{MrNs}\exp(-K_M d_M - K_N d_N)$, B being a constant matrix, and d_M the distance down the Mth guide from the cavity opening to the terminal plane. It is clear that this nondiagonal contribution can be made negligible by suitable choice of the terminal planes, and it will be supposed that this has been done. The above results may then be distilled into the basic formula

$$i_{Ns} = \sum_M \sum_r Y_{Ns\,Mr} v_{Mr} \tag{2.40}$$

or $i = Yv$ in matrix notation. The *admittance matrix* Y itself is given by

$$Y = Z^{\frac{1}{2}} A Z^{\frac{1}{2}} + Y^0. \tag{2.41}$$

Here $Z^{\frac{1}{2}}$ denotes a diagonal matrix with elements $Z_{Mr}^{\frac{1}{2}}$, A the matrix defined by formula (2.38), and Y^0 a matrix all whose elements are zero except diagonal elements with *odd* index s, corresponding to magnetic modes,

$$Y_{NsNs}{}^0 = -jK_{Ns}/k_{Ns}. \tag{2.42}$$

This last assertion follows from the suitable choice of the terminal planes. Because of the symmetry between pairs of normal modes (see (2.28) (3)), it is customary to rewrite A so that only a sum over positive α is involved.

[14] See, for example, O. D. Kellog, *Foundations of Potential Theory* (Verlag. Julius Springer, Berlin, Germany, 1929), p. 246. Details of this, and several of the following results relating to the irrotational part of the field will be found in reference 13.

The matrix elements of Y then become

$$Y_{MrNs}=j\sum_{\alpha}(2\omega Z_{Mr})^{\frac{1}{2}}\frac{\gamma_{\alpha Mr}\gamma_{\alpha Ns}}{\omega_\alpha{}^2-\omega^2}(2\omega Z_{Ns})^{\frac{1}{2}}$$
$$-jK_{Ns}k_{Ns}{}^{-1}\delta_{MN}\delta_{rs}(1-(-1)^s). \quad (2.43)$$

It is clear that Y is a symmetric matrix, and that the elements of Y which relate to allowed modes are imaginary. Equation (2.43) is the analog of the impedance relation derived by Cauer[15] for multiterminal loss-free networks.

For many purposes it is more interesting to express **E** and **H** in the guides in terms of incoming and emitted waves, as in Eq. (2.4), and to determine the properties of the junction by using the ratios of the amplitudes of the emitted to the incident waves. Thus [see (2.4)]

$$b_{Mr}=\sum_N\sum_s U_{MrNs}a_{Ns} \quad (2.44)$$

or symbolically

$$b=Ua.$$

Since $v=a+b$, $i=a-b=Yv$, it follows that

$$U=(1-Y)/(1+Y). \quad (2.45)$$

U is the *collision* or *scattering* matrix: The element U_{MrNs} represents the amplitude and phase of the emitted wave of type Mr due to unit real amplitude incoming wave of type Ns. U is symmetric because Y is symmetric, and the part of U relating to allowed modes can be shown to be unitary because the matrix elements of Y are imaginary for such modes.

The algebraic properties of U, such as symmetry and unitarity, may be derived more directly, without reference to the details of the junction by simply applying the two reciprocity theorems

$$\int_{+T}(\mathbf{E}'\times\mathbf{H}''-\mathbf{E}''\times\mathbf{H}')\cdot\mathbf{n}dS=0 \quad (2.46)$$

$$\int_{+T}(\mathbf{E}'\times\mathbf{H}''^*+\mathbf{E}''\times\mathbf{H}'^*)\cdot\mathbf{n}dS=0 \quad (2.47)$$

to two fields $\mathbf{E}'$, $\mathbf{H}'$, and $\mathbf{E}''$, $\mathbf{H}''$, expressed in terms of incoming and outgoing waves.[3] By using a method applied to nuclear reactions by Wigner,[16] it is possible to derive the frequency dependence too in this more algebraic manner, but the method outlined above, and due to Fränz and to Slater, is simpler, and has more physical content.

The expressions given above for Y and U are the exact analog of corresponding quantities which occur in the theory of nuclear reactions.[6] Certain consequences of these relations have been investigated rather extensively,[17,18] and the corresponding results for microwave junctions could, of course, be obtained immediately. It is not intended to expatiate on these matters here, but it seems worthwhile to devote the next brief section to the statement of two important results which illuminate the nature of the quantities $\gamma_{\alpha Ns}$ and ω_α, apart from having some intrinsic interest.

III. RESONANCE AND AVERAGE CROSS SECTIONS

The significance of the quantities $\gamma_{\alpha Ns}$ and ω_α can best be illustrated by reference to some properties of the cross sections σ_{MrNs} for scattering from one mode Ns into another Mr. σ_{MrNs} is then the ratio of the power of the emitted wave of type Ns to that of the incident wave Mr, and is given simply by $|U_{MrNs}|^2$. Equations (2.45) and (2.43) show that it will generally be an extremely complicated function of frequency ω; it is, however, possible to obtain the frequency dependence in a small but important frequency range $\omega\sim\omega_\beta$, provided certain reasonable conditions are satisfied.

Since the elements of Y all have poles at the points ω_α, one might be led to the opinion that it is only necessary to consider the term with $\alpha=\beta$ in the sum in Eq. (2.43) when evaluating U for $\omega=\omega_\beta$, and probably for ω in this vicinity. The fact that this is not generally so is attributable to the matrix character of Eq. (2.45). However, if the sum of the terms with $\alpha\neq\beta$ is sufficiently small, their total neglect for $\omega\sim\omega_\beta$ will lead to the correct leading term in a matrix expansion of U, and the correction will be of the order of the neglected terms. It will be assumed that this is permissible to begin with.

One thus writes

$$Y\cong Y^0+Y^\beta$$
$$Y_{MrNs}\cong Y_{MrNs}{}^0+jZ_{Mr}{}^{\frac{1}{2}}\frac{\gamma_{\beta Mr}\gamma_{\beta Ns}}{\omega_\beta-\omega}Z_{Ns}{}^{\frac{1}{2}}. \quad (3.1)$$

Y^0 is a diagonal matrix, and Y^β is a matrix of rank one. It is then a matter of simple matrix manipulation to calculate U, and one finds, for instance, (see reference 18 for details) that ($Mr\neq Ns$)

$$U_{MrNs}=\frac{-2j}{\omega_\beta-\omega+\Delta_\beta+j\Gamma_\beta/2}\cdot\frac{Z_{Mr}{}^{\frac{1}{2}}\gamma_{\beta Mr}}{1+jY_{MrMr}{}^0}\cdot\frac{Z_{Ns}{}^{\frac{1}{2}}\gamma_{\beta Ns}}{1+jY_{NsNs}{}^0} \quad (3.2)$$

where

$$\Gamma_\beta=\sum_N\sum_s\Gamma_{\beta Ns}$$
$$=\sum_N\sum_s 2Z_{Ns}\gamma_{\beta Ns}{}^2/(1+Y_{NsNs}{}^{02}) \quad (3.3)$$

$$\Delta_\beta=\sum_N\sum_s Z_{Ns}Y_{NsNs}{}^0\gamma_{\beta Ns}{}^2/(1+Y_{NsNs}{}^{02}). \quad (3.4)$$

[15] W. Cauer, Gött. Nach. N. F. 1 (Fachgruppe I) 1 (1934). In the case of microwave junctions, the non-irrotational part of Y, i.e., the first term of (2.43) was given by Fränz and by Slater (see reference 10).

[16] E. P. Wigner, Phys. Rev. **70**, 606 (1946).

[17] E. P. Wigner, Phys. Rev. **73**, 1002 (1948).

[18] T. Teichmann, Phys. Rev. **77**, 506 (1950).

For the off-diagonal elements one, therefore, has

$$\sigma_{M_\tau N_s}=\frac{\Gamma_{\beta M_\tau}\Gamma_{\beta N_s}}{\Gamma_\beta^2}\cdot\frac{\Gamma_\beta^2/4}{(\omega_\beta+\Delta_\beta-\omega)^2+\Gamma_\beta^2/4} \tag{3.5}$$

plus correction terms of the order

$$\sum_{\alpha\neq\beta} Z_{M_\tau}{}^{\frac{1}{2}}\frac{\gamma_{\alpha M_\tau}\gamma_{\alpha N_s}}{\omega_\alpha-\omega_\beta}Z_{N_s}{}^{\frac{1}{2}}=0(\pi\Gamma/D),$$

where Γ and D are averages of the Γ_α and the frequency spacings $\omega_{\alpha+1}-\omega_\alpha$ in the neighborhood of the frequency considered. If the ratio Γ/D is small enough ($<1/2\pi$), the expansion is permissible, and (3.5) will give the leading term in the cross section. If $\Gamma/D\ll 1$, (3.5) will be a very good approximation to the exact cross section. Provided that ω_β is not in the immediate neighborhood of the cut-off frequencies of the modes $M\tau$ or Ns, the quantities $\Gamma_{\beta N_s}$ and Δ_β will be very slowly varying functions of frequency as compared to $\omega_\beta-\omega$, and $\sigma_{M_\tau N_s}$ will have the form shown in Fig. 2. It is also clear from this figure that the terms *width* for Γ_β, and resonant frequency or *level* for ω_β are the appropriate ones. The $\Gamma_{\beta N_s}$ are known as partial widths, and the frequency independent quantities $\gamma_{\beta N_s}{}^2$ are called reduced widths.

If Γ/D is not small, a series expansion of the matrix quotient (2.45) is not permitted, and the formulas (3.2) and (3.5) may be entirely misleading. It can in fact be shown (reference 18) that under some conditions $\sigma_{M_\tau N_s}$ can have minima at the points ω_α for $\Gamma/D>1$. Such conditions would, however, normally occur only at very high frequencies, or for junctions which are coupled very strongly to the guides, and where the frequency dependence of the $\Gamma_\alpha{}^2$ is comparable to that of the resonance denominator $(\omega_\alpha+\Delta_\alpha-\omega)^2+\Gamma_\alpha{}^2/4$. The normal modes of such a junction under such conditions have mathematical significance, but lose their physical significance as resonances in the cross sections.

It is not proposed to discuss the frequency dependence of the cross sections any further at this time, particularly since more detailed considerations would of necessity require specialized assumptions about the $\gamma_{\alpha N_s}$ and the ω_α, which are probably not justified without reference to cavities of special shapes. It does, however, seem appropriate to conclude this sections by giving the form of the average cross section.

In the case of narrow resonance lines (i.e., $\Gamma/D\ll 1$) it is quite easy to integrate (3.5) over a frequency range including several resonances, and one obtains

$$\sigma_{M_\tau N_s}{}^{(\text{ave})}=\sigma_{M_\tau N_s}{}^{(\max)}\pi\Gamma/2D, \tag{3.6}$$

where

$$\sigma_{M_\tau N_s}{}^{(\max)}=\Gamma_{M_\tau}\Gamma_{N_s}/\Gamma^2.$$

Here the Γ and D are averages over the region considered. This is clearly not correct when Γ/D becomes appreciable, but in this case one is able to show by reference to various special but representative cases (see reference 18) that the expression on the right of (3.6) must be multiplied by an additional function $f(\Gamma/D)$ which has the properties that $f(0)=1$ and that $yf(y)$ tends to 1 as y tends to ∞.

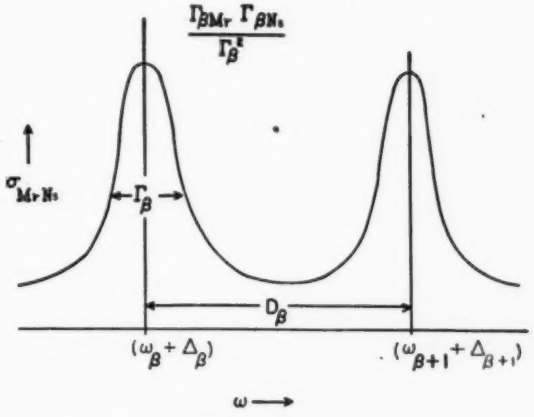


FIG. 2. Form of the resonance cross section.

The quantities $\gamma_{\alpha N_s}{}^2$ and ω_α thus play an entirely analogous role in the theory of microwave junctions to that of the reduced widths (denoted by the same symbol) and the resonant energy levels E_α of the compound nucleus in the theory of nuclear reactions. It is not fruitful to dwell further on this aspect of the analogy, for the reasons stated in Sec. I. The remainder of this paper is thus devoted to the general relations satisfied by the $\gamma_{\alpha N_s}{}^2$ and the ω_α by virtue of the completeness of the wave-guide modes, and of the cavity modes.

IV. COMPLETENESS RELATIONS

The relations to be derived below follow from the fact that the vectors $\mathbf{F}_{N_s}$, $\mathbf{G}_{N_s}$ form a complete set for the expansion of an arbitrary electromagnetic field $\mathbf{E}, \mathbf{H}$ inside the Nth guide, while the vectors $\mathbf{E}_\alpha$, $\mathbf{H}_\alpha$ form a complete set for the expansion of the solenoidal part of $\mathbf{E}, \mathbf{H}$ inside the junction V. The quantities $\gamma_{\alpha N_s}$ may then be regarded as the expansion coefficients with respect to either of these sets of suitably chosen functions.

Completeness in the Guide

The vectors

$$\mathbf{h}_{N_s}=(Z_{N_s}/2\omega)^{\frac{1}{2}}\mathbf{G}_{t,N_s} \tag{4.1}$$

form a complete orthonormal set of vectors across the terminal plane T_N (with respect to transverse vectors lying in this plane). Thus

$$\int_{T_N}\mathbf{h}_{N_s}\cdot\mathbf{h}_{N_r}dT_N=\delta_{rs}. \tag{4.2}$$

Because of the boundary condition (2.25) $\mathbf{H}_\alpha$ is a transverse vector on T_N, and hence may be expanded in terms of the $\mathbf{h}_{N_s}$. By Eq. (2.36)

$$\gamma_{\alpha N_s}=-j\int_{T_N}\mathbf{H}_\alpha\cdot\mathbf{h}_{N_s}dT_N \tag{4.3}$$

and H_α may, therefore, be represented on T_N by the series

$$\mathbf{H}_\alpha = j \sum_s \gamma_{\alpha N s} \mathbf{h}_{N s}. \tag{4.4}$$

The Parseval completeness relation then asserts that

$$\sum_s \gamma_{\alpha N s}^2 = -\int_{T_N} \mathbf{H}_\alpha^2 dT_N. \tag{4.5}$$

(Since $\mathbf{H}_\alpha$ is imaginary the sum is positive.) This relation is exact. The sum on the left is to be extended over all the modes s in the guide, not only over the allowed modes. It is clear from the remarks of Sec. II that if the terminal planes are taken sufficiently far out from what may be regarded as the natural position (i.e., at or near the cavity openings), the cut-off $\gamma_{\alpha N s}$ can be made negligible, because of the exponential decay of the corresponding fields down the guide, and the sum need only be extended over allowed modes. On the other hand, $\mathbf{H}_\alpha$ will then be much more difficult to calculate, or even estimate. Thus one will generally have to compromise, depending on which side of the equation one wants to attach more physical significance to. This problem will, of course, only arise for the lowest mode or modes of the junction; for higher modes it would generally be a matter of prohibitive labor, if at all possible, to calculate $\mathbf{H}_\alpha$ on T_N, and it is necessary to estimate the sum.

A suitable estimate may be obtained as follows: From Eq. (2.30) the average of $\mathbf{H}_\alpha^2$ over V is found to be simply

$$\langle \mathbf{H}_\alpha^2 \rangle_V = -1/2\mu V. \tag{4.6}$$

It will be possible to use this average in place of $\mathbf{H}_\alpha^2$ in the integral on the right of Eq. (4.5) if α is a fairly high mode, whose field oscillates several times over the terminal plane T_N. A rough criterion for the applicability of this approximation is

$$\omega_\alpha (\epsilon\mu T_N)^{\frac{1}{2}} \gg 1. \tag{4.7}$$

This then leads to the estimate

$$\sum_s \gamma_{\alpha N s}^2 = T_N/2\mu V = (T_N/T)(T/2\mu V). \tag{4.8}$$

$(T/2\mu V)$ is a constant of the junction as a whole, while (T_N/T) represents the fractional aperture of the Nth guide.

Such an estimate is not generally possible for the lower modes, for the symmetry properties of the junction and guides may lead to very inequitable distributions of the field over the junction at low normal frequencies. This effect may, however, be averaged out by taking an average of the $\gamma_{\alpha N s}^2$ over several adjoining levels α, and Eq. (4.8) holds for such averages over the entire frequency range. It is possible to obtain an upper limit for the sum on the left of Eq. (4.8) (without any average over α) by assuming that $\mathbf{H}_\alpha$ is the ultimate in unevenness, and piles up in one small region near the Nth guide, which then gives the main contribution to the normalization integral. Supposing that this field is concentrated entirely in a cylinder of base T_N and depth one-quarter wavelength $(\pi/2\omega_\alpha(\epsilon\mu)^{\frac{1}{2}})$, and is sinusoidal in this region, one finds that

$$\sum_s \gamma_{\alpha N s}^2 \leqslant (4\omega_\alpha(\epsilon\mu)^{\frac{1}{2}} V/\pi T)(T/2\mu V). \tag{4.9}$$

(See reference 7 for details of the calculation for the nuclear case.) A maximizing factor $m_\alpha = 4\omega_\alpha(\epsilon\mu)^{\frac{1}{2}} V/\pi T$ thus replaces the factor T_N/T of Eq. (4.8). Using the asymptotic expression

$$\omega_\alpha(\epsilon\mu)^{\frac{1}{2}} \cong (2\pi^2\alpha/V)^{\frac{1}{2}} \tag{4.10}$$

and the well-known inequality relating the volume and surface of a solid[19]

$$36\pi V^2 \leqslant (S+T)^3 \text{ (in this case)} \tag{4.11}$$

one finds

$$m_\alpha \leqslant (32/9\pi^2)^{\frac{1}{2}}(1+g)\alpha^{\frac{1}{2}}, \tag{4.12}$$

where

$$g = S/T \tag{4.13}$$

is the ratio of conducting to aperture surface for the junction. Thus,

$$\sum_s \gamma_{\alpha N s}^2 \leqslant 0.712\alpha^{\frac{1}{2}}(1+g)(T/2\mu V). \tag{4.14}$$

If the geometry of the junctions is well known, it is unnecessary to use the general, and possibly crude inequalities invoked above.

Completeness in the Junction

It is considerably more difficult to obtain a result as precise as the above is (at least in principle) for the sum $\sum_{\alpha \leqslant \lambda} \gamma_{\alpha N s}^2$. The trouble is that the $\gamma_{\alpha N s}$ are not the expansion coefficients with respect to the $\mathbf{H}_\alpha$ of any reasonable function, and, in fact the infinite sum $\sum_{\alpha=1}^{\infty} \gamma_{\alpha N s}^2$ diverges. It is, however, possible to find a function whose expansion coefficients with respect to the $\mathbf{H}_\alpha$ are closely related to the $\gamma_{\alpha N s}$, and which makes it possible to estimate the sum $\sum_{\alpha \leqslant \lambda} \gamma_{\alpha N s}^2$. From the mathematical point of view, the problem is closely related to the determination of the asymptotic form of the $\mathbf{H}_\alpha$ as α tends to ∞.[20,21]

To reach the desired result, one introduces a so-called *comparison* vector $\mathbf{J}_{N s}$ defined as follows: $\mathbf{J}_{N s}$ is zero except in a cylinder with T_N as base, and extending *into*

[19] G. Pólya and G. Szegö, *Isoperimetric Inequalities in Mathematical Physics* (Princeton University Press, Princeton, New Jersey, 1951).

[20] S. Minakshisundaram and A. Pleijel, Can. J. Math. **1**, 242 (1949).

[21] H. Weyl, Bull. Am. Math. Soc. **56**, 115 (1950).

the junction a small distance z, i.e., in the opposite direction to $\mathbf{e}_N$; in this cylinder $\mathbf{J}_{Ns}$ is independent of the displacement from T_N, while as a function of position on T_N it is given simply by

$$\mathbf{J}_{Ns}=\mathbf{h}_{Ns}. \tag{4.15}$$

The $\mathbf{H}_\alpha$ are complete in V with respect vectors which have zero normal component on the surface $S+T$ of V. Since $\mathbf{J}_{Ns}$ is such a vector, one may write

$$\mathbf{J}_{Ns}=j\sum_\alpha X_{\alpha Ns}\mathbf{H}_\alpha, \tag{4.16}$$

where

$$X_{\alpha Ns}=-2\mu j\int_V \mathbf{H}_\alpha\cdot\mathbf{J}_{Ns}dV. \tag{4.17}$$

By the Parseval completeness relation one obtains here

$$\sum_\alpha X_{\alpha Ns}{}^2=2\mu\int_V \mathbf{J}_{Ns}{}^2dV$$
$$=2\mu z, \tag{4.18}$$

using Eq. (4.2). $X_{\alpha Ns}$ may be written in the form

$$X_{\alpha Ns}=2\mu z\gamma_{\alpha Ns}Q_{\alpha Ns}, \tag{4.19}$$

where $Q_{\alpha Ns}$ is a suitably chosen function of α. If z is small, $Q_{\alpha Ns}$ will be almost constant and equal to unity for $\alpha\ll\lambda(z)$, $\lambda(z)$ being approximately equal to the first value of α for which $\mathbf{H}_{t,\alpha}$ has a node in the cylinder of definition of $\mathbf{J}_{Ns}$. $\lambda(z)$ is thus defined by

$$\omega_\lambda(\epsilon\mu)^{\frac{1}{2}}=\pi/2z \tag{4.20}$$

and alternatively, this may be regarded as defining the value of z for which the required effect occurs in terms of (fixed) λ. For $\alpha>\lambda$, $Q_{\alpha Ns}{}^2$ will fall off rapidly in an oscillatory manner as illustrated in Fig. 3. The last three equations then yield the result that

$$\sum_\alpha^\infty \gamma_{\alpha Ns}{}^2Q_{\alpha Ns}{}^2=(\epsilon/\mu)^{\frac{1}{2}}\omega_\lambda/\pi. \tag{4.21}$$

If the exact form of the $\mathbf{H}_\alpha$ were known, one could calculate the $Q_{\alpha Ns}$ and the sum on the right of Eq. (4.21) would be precisely defined. In this event, however, the $\gamma_{\alpha Ns}$ could also be exactly calculated, and there would be no need to go through this indirect procedure to obtain the sum. It is just because of a lack of such precise knowledge, particularly for the higher modes, $\alpha\gg1$, that it is necessary to embark on the more formal, more general, but less explicit approach developed above. One should, therefore, not be excessively timid about making simple but suitable approximations for $Q_{\alpha Ns}{}^2$. The one which suggests itself is to take

$$\begin{aligned} Q_{\alpha Ns}{}^2&=1 \quad \text{for} \quad \alpha\leqslant\lambda \\ Q_{\alpha Ns}{}^2&=0 \quad \text{for} \quad \alpha>\lambda, \end{aligned} \tag{4.22}$$

as shown in Fig. 3. In the nuclear case (see reference 7)

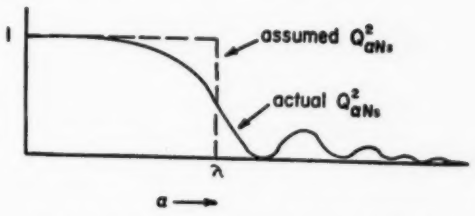


FIG. 3. Actual and assumed forms of $Q_{\alpha Ns}{}^2$.

this leads to a formula which is asymptotically correct, for large α, and one may, therefore, have a fair amount of confidence in it in the present application. One therefore finds

$$\sum_{\alpha\leqslant\lambda}\gamma_{\alpha Ns}{}^2\cong(\epsilon/\mu)^{\frac{1}{2}}\omega_\lambda/\pi, \tag{4.23}$$

which is the required formula. It is more instructive, however, to derive from this relation the behavior of $\gamma_{\alpha Ns}{}^2$ as a function of α for high α. It is evident that this is given simply by

$$\langle\gamma_{\alpha Ns}{}^2\rangle\cong(\epsilon/\mu)^{\frac{1}{2}}D_\alpha/\pi. \tag{4.24}$$

D_α denotes the average spacing of the proper frequencies (levels) in a region round ω_α containing at least several modes, and $\langle\gamma_{\alpha Ns}{}^2\rangle$ denotes the average of $\gamma_{\alpha Ns}{}^2$ in the same region.

One may apply similar arguments to two comparison vectors $\mathbf{J}_{Ns}$ and $\mathbf{J}_{Mr}$, not necessarily defined in the same aperture. In this case

$$\sum_\alpha X_{\alpha Mr}X_{\alpha Ns}=2\mu\int_V \mathbf{J}_{Mr}\cdot\mathbf{J}_{Ns}dV$$
$$=2\mu z\delta_{MN}\delta_{rs}, \tag{4.25}$$

and hence, the $\gamma_{\alpha Ns}$ satisfy the more general asymptotic relations

$$\langle\gamma_{\alpha Mr}\gamma_{\alpha Ns}\rangle\cong(\epsilon/\mu)^{\frac{1}{2}}D_\alpha\pi^{-1}\delta_{MN}\delta_{rs}. \tag{4.26}$$

For $M\neq N$, or $r\neq s$ this is a special case of Weyl's *incoherency* relation (see reference 21).

The analogous relations in the theory of nuclear reactions are known as *sum rules*, since they determine to a large extent the total transition probability from a state α of the compound system to the residual states s (Eq. (4.8)), and conversely the total transition probability from an incident state s to a set of states α of the compound system. The term *sum rule* is also applicable in this, the electromagnetic case, provided that one bears in mind that the sums are now of coupling coefficients (the $\gamma_{\alpha Ns}{}^2$), and that they now determine the actual transfer of electromagnetic energy between the guide modes and those of the junction, rather than the transition probability between a compound system and its possible disintegration products.

V. APPLICATIONS

The results of the last section make it possible to estimate the part played by the higher modes of the

guides and of the junction on the admittance matrix Y, and hence on the transfer of electromagnetic power and signals through the junction.

Combining Eq. (4.8) in its second interpretation with Eq. (4.24), it immediately becomes evident that only a finite number of wave-guide modes can contribute appreciably to the sum in (4.8). Introducing the average lowest characteristic frequency of the junction, ω_c, given by

$$\omega_c=\pi(\epsilon\mu)^{-\frac{1}{2}}V^{-\frac{1}{3}}, \tag{5.1}$$

this number $n_{N\alpha}$ is found to be

$$n_{N\alpha}=(T_N/2V^{\frac{1}{3}})(\omega_c/D_\alpha). \tag{5.2}$$

The first factor on the right of Eq. (5.2) represents the influence of the coupling between the guides and the junction (purely geometrically), while the second factor depends only on the shape of the junction and on the region of excitation (near α) considered. $n_{N\alpha}$ may thus be regarded as the number of wave-guide modes required in the Nth guide to excite the junction uniformly over a unit range of frequency in the region near ω_α (and including at least several resonances), *assuming that all these guide modes contribute equally.* The mathematical counterpart of this picture is the assertion that the sum $\sum_s \gamma_{\alpha N s}^2$ converges because only $n_{N\alpha}$ of the $\gamma_{\alpha N s}^2$ are finite (and more or less of the same order of magnitude), the remainder being zero. This is to be contrasted with the usual interpretation that the convergence is due to uniform decrease in the $\gamma_{\alpha N s}^2$ with increasing s. This latter view is generally in better agreement with the actual situation, but nevertheless does not give one the same grip on it that is obtained by the formal introduction of the number $n_{N\alpha}$. Formula (5.2) displays all the qualitative features that one might expect on physical grounds: $n_{N\alpha}$ increases with the density of junction modes (i.e., decreases with increasing mode spacing D_α), and also increases with increasing aperture (guide cross section) T_N.

In order to estimate the total effect on the admittance matrix of the higher junction modes, one starts with Eq. (4.26) in the form

$$\gamma_{\alpha M r}\gamma_{\alpha N s}\cong\frac{1}{\pi}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\frac{d\omega_\alpha}{d\alpha}\delta_{MN}\delta_{rs}, \tag{5.3}$$

which follows from Eq. (4.23) and (4.25) if the modes α are supposed closely spaced. If this is the case for $\alpha>\beta$, then one may write

$$2\omega\sum_{\alpha\gtrless\beta}^{\infty}\frac{\gamma_{\alpha M r}\gamma_{\alpha N s}}{\omega_\alpha^2-\omega^2}\cong\frac{2\omega}{\pi}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\int_{\omega\beta}^{\infty}\frac{d\omega_\alpha}{\omega_\alpha^2-\omega^2}\delta_{MN}\delta_{rs}$$

$$=\frac{1}{\pi}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\log\left(\frac{\omega_\beta+\omega}{\omega_\beta-\omega}\right)\delta_{MN}\delta_{rs}, \tag{5.4}$$

in which case the contribution of the terms $\omega_\alpha>\omega_\beta$ to $Y_{M r N s}$ becomes

$$Z_{N s}\delta_{MN}\delta_{rs}\cdot\frac{1}{\pi}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\log\left(\frac{\omega_\beta+\omega}{\omega_\beta-\omega}\right). \tag{5.5}$$

The above calculation presupposes that $\omega>\omega_\beta$. In particular, if $\omega\sim\omega_1$, the lowest mode of the junction, and it is possible to replace the rest of the sum for the elements of Y by an integral, then the aggregate of nonresonant terms contribute the amount

$$Z_{N s}\delta_{MN}\delta_{rs}\cdot\frac{1}{\pi}\left(\frac{\epsilon}{\mu}\right)^{\frac{1}{2}}\log\left(1+2\frac{\omega_1}{D_1}\right). \tag{5.6}$$

Reference to Sec. III shows that this term will manifest itself mainly in an alteration of the width of the resonance cross section (3.5). (Because of its diagonal character it may be treated simply as an addition to Y^0.)

ACKNOWLEDGMENTS

I am indebted to Professors R. H. Dicke and E. P. Wigner for making available to me some of their unpublished work, and for helpful and critical comments. I should also like to thank Drs. E. T. Jaynes, M. Moshinsky, and G. S. Newell for suggestive discussions.

This work was assisted in part by the AEC and the Higgins Fund.

Methods of Measuring the Properties of Ionized Gases at High Frequencies. I. Measurements of Q*

SANBORN C. BROWN AND DAVID J. ROSE†
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts
(Received January 24, 1952)

Experimental methods are given for determining the Q of both high and low Q resonant cavities at microwave frequencies. The emphasis is placed on the practical measurements necessary in determining the properties of ionized gases in the 3- and 10-centimeter wavelength range.

I. INTRODUCTION

THE phenomena associated with electrical discharges in gases at microwave frequencies are in general much simpler than the corresponding behavior of gases under the action of a dc field. On the other hand, microwave techniques are not as well known and may therefore discourage workers from entering the field. Experience has shown that readily available texts on ultra high frequency measurements do not include the particular calculations of the experimental parameters in a convenient form for gas discharge measurements and the present series of papers has been prepared to fill this need.

For the sake of convenience, the range of frequencies is chosen so that cavity resonators may be employed which are relatively simple to construct and use, and where power sources and measuring apparatus are available. One of the important parameters which control the physical behavior of the gas discharge is the value of the electric field in which the electrons are accelerated. Measurements of the field generally involve the prior determination of the power absorbed in a resonant cavity and the determination of the characteristic of the cavity known as its "Q," which is the ratio of the energy stored by the cavity to the power dissipated in the cavity and certain associated components per radian of rf field. In the study of ionization processes, breakdown, and low current-density discharge phenomena, high Q devices may be used. For high current-density discharges, low Q devices are necessarily used. Explicit use of the Q in calculating electric fields is mainly confined to high Q devices; thus this paper is concerned principally with this case. However, certain calculations apply to low Q cavities, and we shall present these also.

In the problem of calculating the Q of resonant cavities used in gas discharge measurements, the necessary accuracy is not achieved if one neglects the series losses in the coupling circuit between cavity and transmission line. Since the total loading of the cavity determines the width of the resonance curve, a quantity known as the loaded Q gives the best coverage of the parameters. For gas discharge work, the Q with the series loss included is calculated in terms of the loaded Q. This method extends the work of Lawson[1] in which the calculations neglected the series loss, and differs from that developed by Slater[2] which included the series loss but was calculated in terms of an external Q.

II. THE LUMPED CIRCUIT REPRESENTATION OF A CAVITY

A cavity may be represented as an element in a microwave circuit as shown in the circuits of Fig. 1. The cavity system consists of a cavity and a part of the input line to which the cavity is coupled by loops through vacuum tight glass seals. Our calculations apply explicitly to the usual case in which the cavity terminates the line, but certain cases in which additional elements are placed beyond the cavity are susceptible to the same treatment. The cavity can be represented by an impedance Z' at TT' which is transformed by the loop circuit to the impedance Z at a reference plane in the input line PP'. We will develop methods for determining the impedance Z, the Q, and the resonant wavelength λ_0 of the cavity system from measurements of the voltage distribution on the line terminated by the cavity. The measurements consist of determining the position of the voltage minima d and the voltage standing wave ratio in decibels, as functions of the wavelength λ. The representation of the cavity and coupling system by Z at PP' will only be valid as long

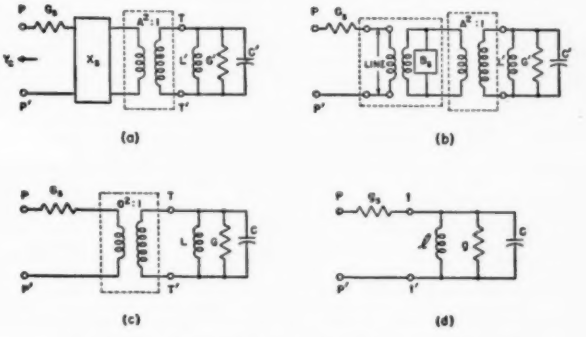


FIG. 1. Equivalent circuits for a resonant cavity.

* This work has been supported in part by the Signal Corps, Air Materiel Command, and ONR.
† Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] G. B. Collins, *Microwave Magnetrons* (McGraw-Hill Book Company, Inc., New York, 1948), Sec. 18.5.
[2] J. C. Slater, Revs. Modern Phys. 18, 473 (1946).

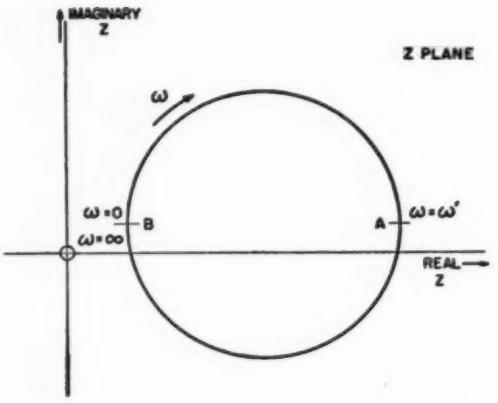


FIG. 2. Approximate impedance diagram for circuit of Fig. 1(a).

as we stay outside of the reference plane, that is, to the left of PP' of Fig. 1.

In order to be complete, the equivalent circuit should contain sufficient elements to account for all phenomena represented by the cavity. Practically, such a circuit is very difficult to work with in detail. First, therefore, we will concern ourselves with the general characteristics of a fairly complete equivalent circuit, and then, following certain restrictions, its simplification to usable form. Let us first consider the circuit of Fig. 1(a), where measurements are made at the points PP' on the line. The resonant cavity, across TT', can be characterized by parallel reactance L' and C', and conductance G', connected to the line by a transformer of admittance ratio $A^2:1$. This representation of a cavity with an isolated resonance can be justified by fairly rigorous methods.[2] In order that the representation be correct, it is necessary that the reactive elements L' and C' refer only to the cavity in question. For example, the coupling of a second cavity onto the first cannot generally be treated mathematically in this manner. However, the conductance G' may well include the effect of nonresonant elements coupled to the cavity as outputs of it. To this extent, then, the cavity need not terminate the transmission line. The reactance of the loop is represented by the lossless network X_s which is often treated as a series inductance; and the losses in the loop, seals, and line back to PP' are represented by the conductance G_s. In the following analysis, it is assumed that the line of characteristic admittance Y_c is matched at PP' looking toward the generator.

Following the general method given by Slater,[3] one may demonstrate the characteristics of this circuit. For the sake of discussion, the network X_s is replaced by a series inductance L_s. If the series reactance $jL_s\omega$ is a much more slowly varying function of frequency than the cavity reactance, the impedance Z at PP' is approximately a circle in the complex impedance plane as shown in Fig. 2. We get this circle by splitting the complex function Z, derived for circuit 1(a), into the real and imaginary parts and by elimination of ω from these two parts. Point A corresponds to the cavity resonance at $\omega=\omega'$; point B, nominally at $\omega=0$ and $\omega=\infty$, may be approached very closely, a few percent off resonance; its position is at $Z=1/G_s+jL_s\omega'$, in so far as $\omega\approx\omega'$ throughout the range of interest.

A simple analysis near resonance is possible only if the energy storage in the loop circuit is negligible compared with that in the cavity at resonance. If the loop is nonresonant, an approximately equivalent criterion is that the ratio of energy stored off resonance to energy stored on resonance is negligible; or equivalently, that the cavity resonance curve (described in detail below) shows no signs of additional resonant frequencies near the main one. Under these assumptions, the loop elements are a small perturbation, and may be taken into account in the following manner. The lossless network X_s is replaced by a length of line, a transformer, and a susceptance B_s, which is always possible, giving the circuit of Fig. 1(b). The elements are then transformed to those of Fig. 1(c), in which the loop reactance, which is the principal complication to further analysis, has been combined with the other cavity parameters. This simplified circuit is the basis for all subsequent analysis. The new cavity parameters are L, G, and C; and the new transformer admittance ratio, obtained from combining the two transformers, is $a^2:1$. It is seen that the principal effect of the loop is to shift the resonant frequency to ω_0, where $\omega_0=1/(LC)^{\frac{1}{2}}$, and to change the effective line length between the cavity and PP'. The effective line length is defined as the electrical distance between TT', the points at which the cavity alone could be represented by a single impedance and PP', the points at which the cavity plus loop circuit may be represented by a single impedance. Thus if only the coupling loop were changed, TT' would stay fixed but PP' would change. The points PP' are determined on a lossless line only to integral multiples of $\lambda/2$.

For many purposes, this circuit may again be replaced by that of Fig. 1(d), wherein the line admittance is normalized to unity. Here

$$g_s=G_s/Y_c,\quad l=LY_c/a^2,\quad g=a^2G/Y_c,\quad \text{and}\quad c=Ca^2/Y_c. \tag{1}$$

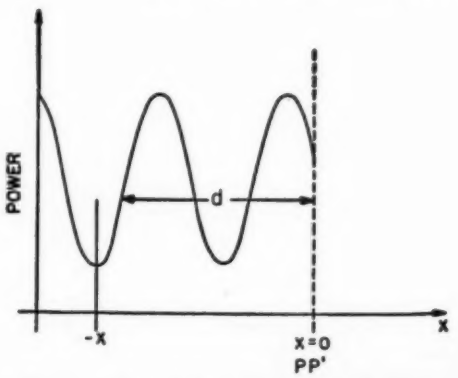


FIG. 3. Superposition of the incident and reflected waves.

[3] J. C. Slater, reference 2, pp. 483–489.

At resonance, the susceptance of the parallel $l-c$ combination is zero. It is now convenient to define two quantities, the resonant conductance across PP' as g_0, where

$$g_0=g/(1+g/g_e), \tag{2}$$

and a dimensionless quantity β, akin to susceptance, given by

$$\beta=(c/l)^{\frac{1}{2}}=a^2(C/L)^{\frac{1}{2}}/Y_c. \tag{3}$$

The line is terminated at PP' with the normalized impedance

$$z=1/g_e+1/(g+jb), \tag{4}$$

where

$$b=\omega c-1/\omega l. \tag{5}$$

At resonance $b\rightarrow 0$ and $z_0=1/g_e+1/g=1/g_0$. We call the cavity system overcoupled if $1/g_0>1$ and undercoupled if $1/g_0<1$.

Let us now determine the relations between the experimentally measured position of the voltage minima and the voltage standing wave ratio, and the cavity parameters enumerated above. Assume a voltage wave $A=A_0\exp j(\omega t-2\pi x/\lambda_g)$ travelling from the generator along the positive x axis toward the cavity represented by the normalized impedance z at PP'. Here, λ_g is the guide wavelength and t is the time. The position $x=0$ can be chosen at any fixed point on the line; it is most convenient to choose it at PP'. There is also a reflected wave $B=B_0\exp j(\omega t+2\pi x/\lambda_g)$. A, A_0, B, and B_0 are all complex quantities. The voltage at any point on the line at any time is the sum of these two waves. A probe inserted into a slotted line and connected to a proper receiver measures a quantity proportional to the time average of the power at that point, that is proportional to the absolute magnitude of the square of the voltage $A+B$. The resulting quantity varies down the line as a $\sin^2$ curve displaced above the axis, as shown in Fig. 3.

The complex ratio of incident to reflected waves is the reflection coefficient Γ, which is seen to be

$$\Gamma=B/A=(B_0/A_0)\exp(j4\pi x/\lambda_g)=\rho\exp(j\phi). \tag{6}$$

Here, $\rho=|\Gamma|$ is the modulus of the reflection coefficient, a quantity which is always real and positive; ϕ is the phase angle. The standing wave ratio R_V, that is the ratio of the maximum to minimum voltage on the line, is seen to be

$$R_V=(1+\rho)/(1-\rho). \tag{7}$$

As mentioned above, relative power is usually the quantity measured. The decibel power standing wave ratio R is given by

$$R=20\log_{10}R_V. \tag{8}$$

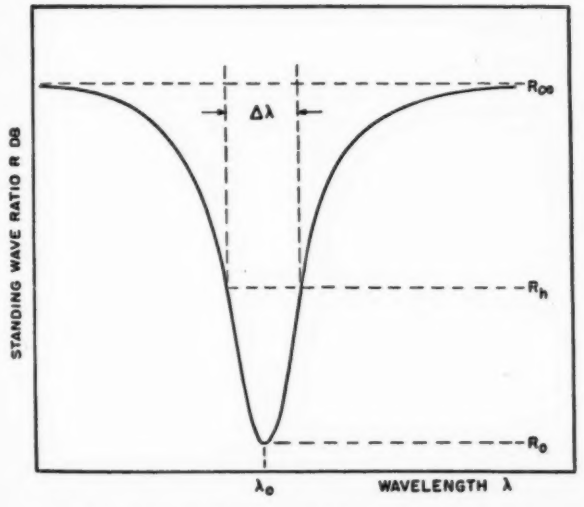


FIG. 4. Typical experimental resonance curve.

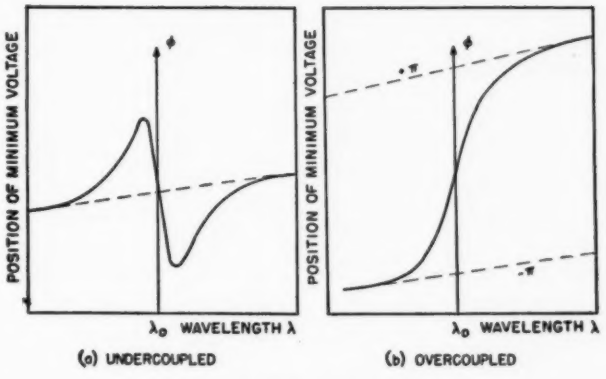


FIG. 5. Typical phase curves.

The experimental procedure consists in setting the wavelength, measuring the standing wave ratio R in db, and also measuring the position d of a voltage minimum. The curve of Fig. 4 is plotted point by point. The ordinates are found by varying λ and setting the probe for each λ value first at a minimum and then at a maximum. With the probe at a minimum, the input attenuation of the receiver is set at a given value. With the probe at a maximum we bring the signal back to the same level by increasing the input attenuation. The ratio of these two attenuator settings gives R. A phase curve which will resemble either Fig. 5(a) or 5(b), depending on whether the cavity is undercoupled or overcoupled, is obtained by plotting the position of the voltage minimum. The asymptotes of Fig. 5 are the actual positions of the points PP', determined within multiples of $\lambda_g/2$. Occasionally the resonance curve is too broad to enable one to reach these asymptotes conveniently, and yet it may be desirable to determine them. In such a case, the cavity should be shorted out in its high field region with a substantial metallic conductor. The resonance is then far removed from ω_0, and the position of the standing wave minimum follows the asymptotes through the region near ω_0. The db standing wave ratio so obtained with this arrangement is R_∞ (Fig. 4).

Before performing the specific algebraic manipulations involved in finding the cavity Q, it remains to

relate the reflection coefficient Γ to the cavity impedance Z. It can be shown in general that the reflection coefficient at any point is given by

$$\Gamma=(z-1)/(z+1)=(1-y)/(1+y), \tag{9}$$

where z is the normalized impedance at the point in question on the line; $y=1/z$ is the corresponding normalized admittance. As may be seen from Eq. (6), the phase angle ϕ of Γ varies with position on the line; the associated impedance z varies correspondingly. Henceforth, the calculations will deal entirely with conditions at the point PP'. It is understood that Γ and ϕ refer to that point. Since z at PP' is given by Eq. (4), Γ and ϕ at PP', and ρ, R_V, and R are all determined.

One can obtain a geometrical representation of Eq. (9), similar to Fig. 2, when the series impedance is neglected, by splitting it into real and imaginary parts and eliminating the reactive element b in the admittance. The Γ circles are the conductance circles of a Smith[4] impedance chart and the tip of the Γ vector determines the susceptance circle. Thus a Smith chart yields ρ and ϕ according to Eq. (9) for every given y or z at PP'. This is shown in Fig. 6.

At PP', Eqs. (4), (6), and (9) can be combined to yield

$$\rho=\left\{4b^2+\left[g^2\left(\frac{1}{g_0^2}-1\right)+b^2\left(\frac{1}{g_s^2}-1\right)\right]^2\right\}^{\frac{1}{2}}\Big/\left[g^2\left(\frac{1}{g_0}+1\right)^2+b^2\left(\frac{1}{g_s}+1\right)^2\right], \tag{10}$$

$$\tan\phi=-2b/\{[g^2(1/g_0^2)-1]+b^2[(1/g_s^2)-1]\}. \tag{11}$$

Insertion of the resonance impedance $z=1/g_0$ (or $b=0$) gives

$$\begin{array}{lll}\text{overcoupled} & \rho_0=(1-g_0)/(1+g_0) & \phi_0=0\\ \text{undercoupled} & \rho_0=(g_0-1)/(g_0+1) & \phi_0=\pi.\end{array} \tag{12}$$

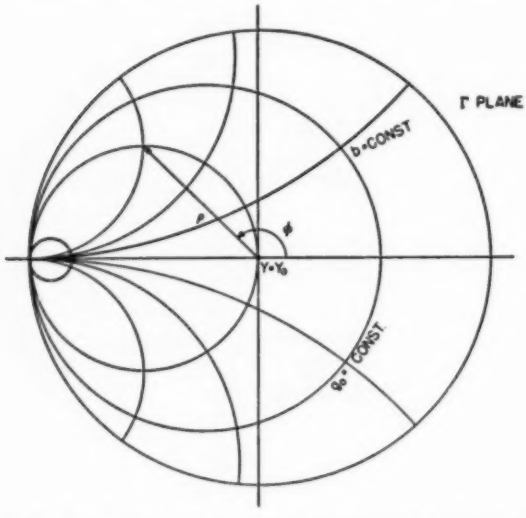


FIG. 6. Smith chart. Circles of constant g_0 and b in the Γ plane.

[4] P. H. Smith, Electronics **12**, 29 (1939) and **17**, 130 (1944).

From Fig. 3, and the definition of ϕ, it is seen that the position d of the standing wave minimum is given by

$$d=(2k-1)\lambda_g/4+\phi\lambda_g/4\pi, \tag{13}$$

where k is an integer which indicates that the kth minimum to the left of PP' is being measured. This variation gives the curves of Fig. 5. The slope of the dotted asymptotes is due to the first term in Eq. (13), as λ_g increases with λ. In the undercoupled case, ϕ has the value π both at resonance and far from resonance. In the overcoupled case, ϕ is zero at resonance, $-\pi$ at the lower asymptote, and $+\pi$ at the upper one. Thus

$$\begin{array}{lll}\text{overcoupled}\begin{cases}\text{upper asymptote} & d=2k\lambda_g/4\\ \text{resonant line} & d=(2k-1)\lambda_g/4\\ \text{lower asymptote} & d=(2k-2)\lambda_g/4\end{cases}\\ \text{undercoupled} \quad \text{asymptote} & d=2k\lambda_g/4.\end{array} \tag{14}$$

One further convention will be mentioned. Although R_V is always greater than unity and therefore R is always greater than zero, it is useful in order to distinguish between the undercoupled and overcoupled cases to define formally $R<0$ for the undercoupled and $R>0$ for the overcoupled cavity systems.

III. THE MEASUREMENT OF Q

The quantity Q is defined as

$$Q=\omega_0 U/P, \tag{15}$$

where U is the energy stored, and P the energy dissipated per second. It may be noted that the concept of a cavity Q is valid only if the stored energy takes many cycles to decay after excitation is removed; i.e., only if $Q\gg 1$. Various kinds of Q's may be discussed, depending on where, in a particular problem, it is convenient to consider the energy to be dissipated. The Q of principal interest in this paper is that obtained by considering the cavity represented in Fig. 1(d) to be excited externally, and calculating the ratio of energy stored in the cavity to that dissipated in g_s and g. This ratio determines the unloaded Q, Q_U, which is, at resonance,

$$Q_U=(c/l)^{\frac{1}{2}}/[g(1+g/g_s)]=\beta(g_s-g_0)^2/g_s^2 g_0. \tag{16}$$

Since the quantities g_s and g_0 can be expressed directly in terms of the resonant db standing wave ratio R_0 and the off-resonance ratio R_∞, the calculations of β and Q_U are closely allied. Two methods are now presented for determining Q_U.

Resonance Curve Method

This method is applicable if the resonance curve is sufficiently sharp so that a certain width $\Delta\lambda$ as shown in Fig. 4 can conveniently be measured on it. In order that this width may be measured where the curve is, in general, most accurately known, the calculation of Q_U is broken into two parts.

It is convenient to define a loaded Q, Q_L, in terms of the ratio of energy stored and energy dissipated through

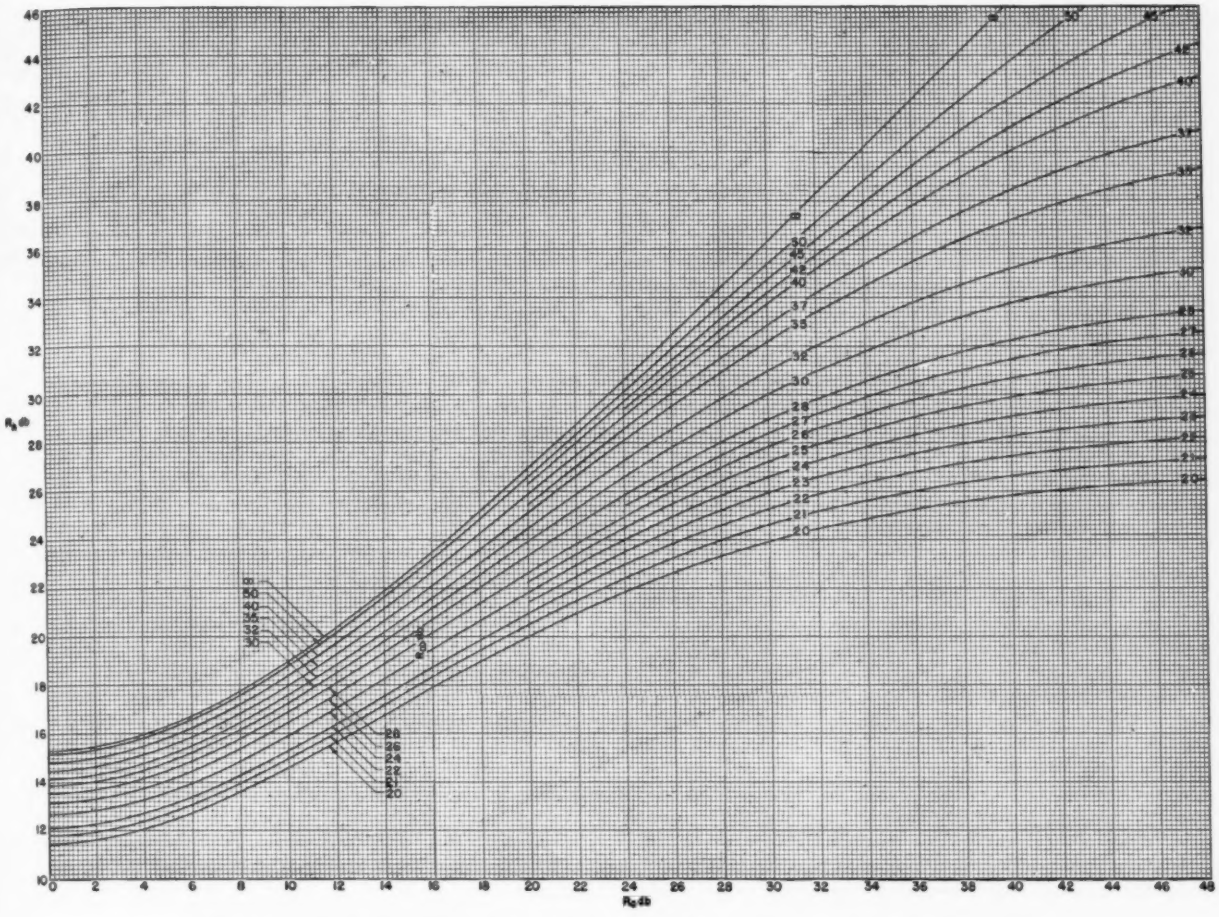


FIG. 7. Standing wave ratio R_h db as a function of the standing wave ratio on resonance.

the parallel conductances g and $g_s/(g_s+1)$, when the cavity and line are viewed from the terminals tt'. Then in terms of the quantities g_s and g_0 given by Eq. (2),

$$Q_L=\beta(g_s-g_0)(g_s+1)/g_s^2(g_0+1)=\lambda_0/\Delta\lambda. \quad (17)$$

This relation also defines the width $\Delta\lambda$ to be measured. Now the $l-c$ susceptance at tt' is

$$jb=j\beta(\omega/\omega_0-\omega_0/\omega) \quad (18)$$

and at $\Delta\lambda/2$ off resonance, the quantity $|\omega/\omega_0-\omega_0/\omega|$ is equal to $\Delta\lambda/\lambda_0$; at these points, the susceptance is equal to the conductance across tt'. Under these conditions, the impedance z_h at PP' is

$$z_h=(1/g_s)+1/\{g+j[g_s^2(g_0+1)/(g_s-g_0)(g_s+1)]\}. \quad (19)$$

The reflection coefficient Γ for the circuit of Fig. 1(d) is given by Eq. (9); thus, using Eq. (10),

$$\rho_h=(g_0^2g_s^2+g_0^2+g_s^2-4g_0g_s+1)^{\frac{1}{2}}/(g_s+1)(g_0+1); \quad (20)$$

R_h is then found from Eqs. (7) and (8). Measurements of the db ratios R_0 and R_∞ yield values of g_0 and g_s. The value of R_h so determined is that at which the width $\Delta\lambda$ is to be measured. These results may be put into the form of Fig. 7 where curves of R_h *vs* R_0 are plotted for various values of R_∞. These curves are identical for positive and negative values of R_0.

The value of Q_L is then determined in the following manner. The ratios R_0 and R_∞ and the resonant wavelength λ_0, are found from the experimental resonance curve as in Fig. 4; then from Fig. 7, the value of R_h is determined. The width of the resonance curve at $R=R_h$ is the required quantity $\Delta\lambda$, and Q_L is found from Eq. (17).

From Eqs. (16) and (17), the unloaded Q is seen to be

$$Q_U=(g_s-g_0)(g_0+1)Q_L/g_0(g_s+1). \quad (21)$$

The required ratio Q_U/Q_L may thus be plotted *vs* R_0 for various values of R_∞. It is shown in Fig. 8. Comparison of the phase curve with Fig. 5 will determine the proper sign of R_0.

The quantity β is most conveniently calculated from Eq. (16) as

$$\beta/Q_L=g_s^2(g_0+1)/(g_s-g_0)(g_s+1). \quad (22)$$

The ratio β/Q_L *vs* R_0 is plotted for various values of R_∞ in Fig. 9; then β is determined from these curves and the value of Q_L. If the series loss can be neglected, the calculations are greatly simplified, for the ratio R_∞ is

essentially infinite. In such a case

$$Q_L=\beta/(g_0+1) \tag{17a}$$

and

$$Q_U=(g_0+1)Q_L/g_0. \tag{21a}$$

Phase Curve Method

When β is quite low, it is often difficult to perform the measurements indicated above. If, for example, β is 5, the required wavelength separation $\Delta\lambda$ is enormous and may easily fall outside the wavelength range of the oscillator used. In this case, however, it is sufficient to measure the slope of the phase curve and the standing-wave ratio on resonance. It will turn out that $g\ll 1$, so that it is permissible near resonance to neglect the series resistance $1/g_s$ and set $g_0=g$. If desired, the value of g_s may be found by the method used for determining the reference points PP'. The circuit takes the form of Fig. 10.

In the vicinity of resonance, the susceptance b, Eq. (18), will be much smaller than g. It is then permissible to neglect all second-order terms in subsequent manipulations. Near resonance, the reflection coefficient Γ is then

$$\Gamma=(1-g-jb)/(1+g+jb)\approx[(1-g^2)-j2b]/(1+g)^2. \tag{23}$$

The small phase angle ϕ is very closely

$$\phi=2b/(g^2-1). \tag{24}$$

The slope of the phase curve $d(d)/d\lambda$ from Eq. (13) is

$$d(d)/d\lambda=[(2k-1)/4]d\lambda_g/d\lambda+(\lambda_g/4\pi)d\phi/d\lambda+(\phi/4\pi)d\lambda_g/d\lambda. \tag{25}$$

This derivative is to be evaluated at $\phi=0$ (minimum position at resonance), $\lambda=\lambda_0$, $\lambda_g=\lambda_{g0}$, using the relation

$$d\lambda_g/d\lambda=(\lambda_g/\lambda)^3.$$

There results

$$\left.\frac{d(d)}{d\lambda}\right|_{\phi=0}=\frac{2k-1}{4}\left(\frac{\lambda_{g0}}{\lambda_0}\right)^3+\frac{\lambda_{g0}}{4\pi}\left.\frac{d\phi}{db}\frac{db}{d\lambda}\right|_{\phi=0}. \tag{26}$$

But $d\phi/db|_{\phi=0}$ is given by Eq. (24), and $db/d\lambda|_{\phi=0}$ is simply $-2\beta/\lambda_0$ by definition of b (Eq. (18)). With these insertions,

$$\left.\frac{d(d)}{d\lambda}\right|_{\phi=0}=\frac{2k-1}{4}\left(\frac{\lambda_{g0}}{\lambda_0}\right)^3+\frac{\beta\lambda_{g0}}{\pi(1-g^2)\lambda_0}, \tag{27}$$

whence

$$\beta=\pi(1-g^2)\left[\left.\frac{\lambda_0 d(d)}{\lambda_{g0}\,d\lambda}\right|_{\phi=0}-\frac{2k-1}{4}\left(\frac{\lambda_{g0}}{\lambda_0}\right)^2\right]. \tag{28}$$

Here, g is the reciprocal of R_{V0}, the voltage standing-

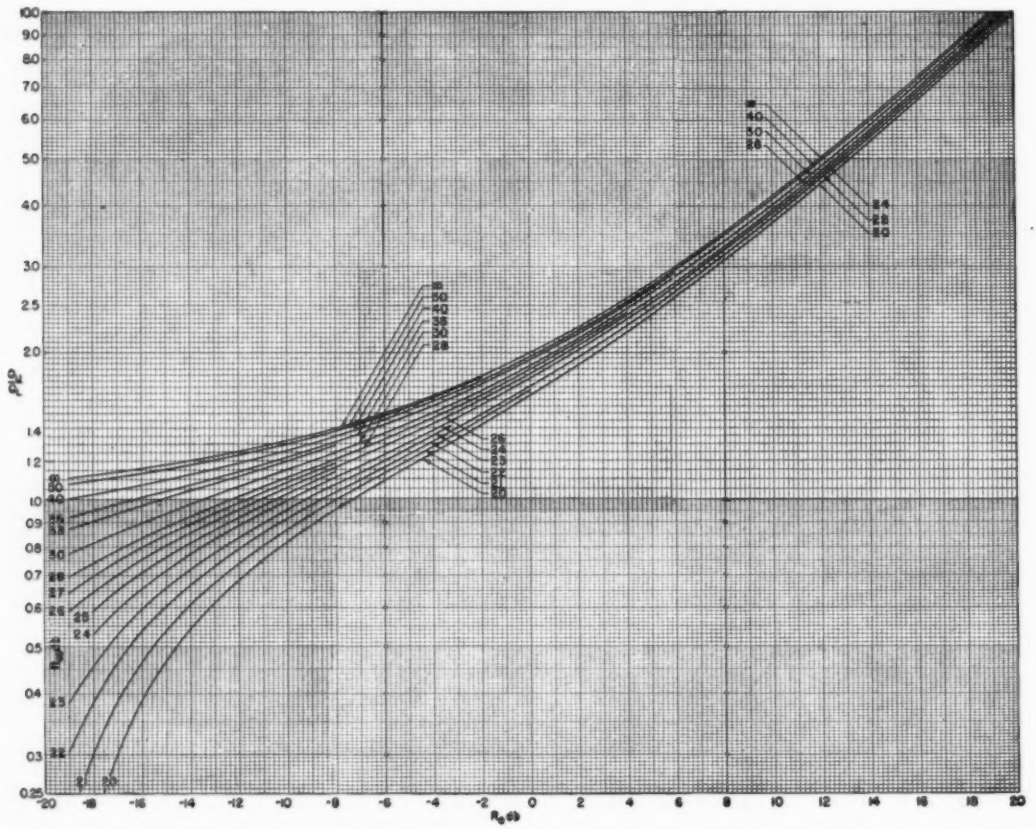


FIG. 8. Ratio of unloaded to loaded Q's plotted as a function of the standing wave ratio on resonance.

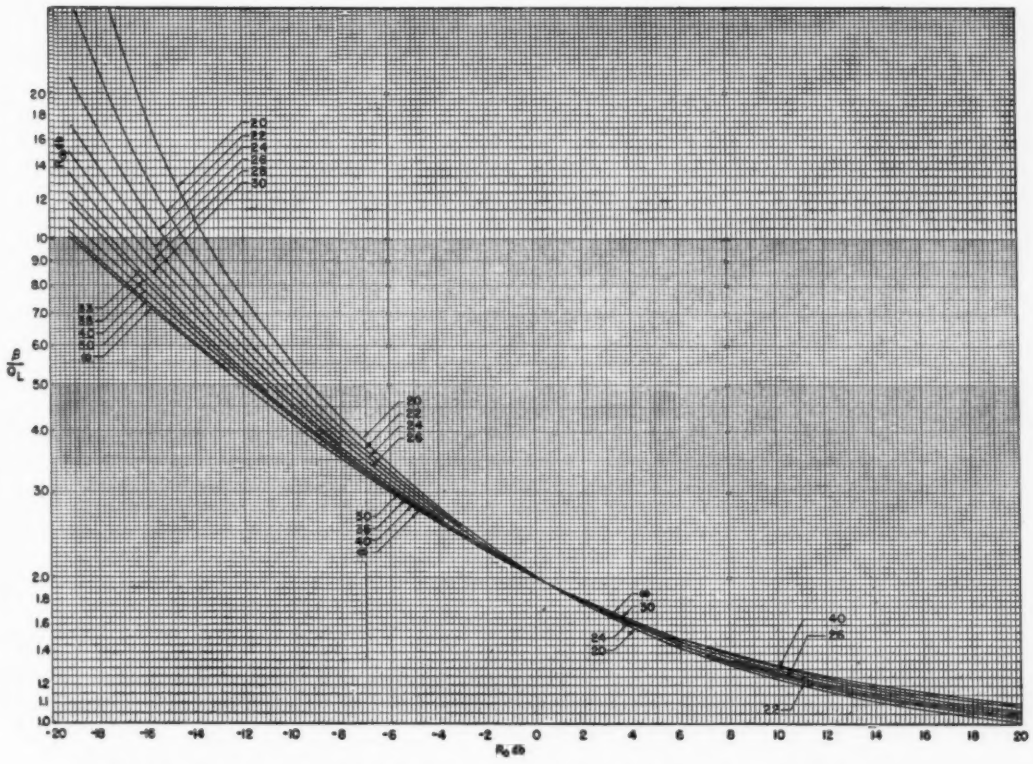


FIG. 9. Ratio of β to the loaded Q plotted as a function of the standing wave ratio on resonance.

wave ratio on resonance; the $(1-g^2)$ term is usually a minor correction when β is small; $d(d)/d\lambda|_{\phi=0}$ is simply the slope of the phase curve at resonance; the factor $(2k-1)/4$ can usually be found by measuring the physical distance in guide wavelengths between the cavity and the point of measurement. The unloaded Q is given in this case by

$$Q_U=\beta/g. \tag{29}$$

High Frequency Bridge Method

At somewhat lower frequencies—of the order of 100 megacycles per second—standing waves still exist on the transmission line and must therefore be taken into account. However, the use of slotted sections is physically impractical, and it becomes desirable to use the high frequency impedance bridge.[5] This instrument is employed commercially in high frequency power and impedance measuring equipment. Essentially it consists of two bridge circuits constructed so that one of them is sensitive to power incident on the load and the other to power reflected from the load. Together they indicate the true absorbed power, and indirectly the absolute value of the reflection coefficient and standing wave ratio of the load.

Since the ratio of reflected power P_r to incident power P_i can be read, it is, in principle, possible to plot a resonance curve for the cavity for

$$|\Gamma|=\rho=(P_r/P_i)^{\frac{1}{2}}, \tag{30}$$

and R_V and R are determined from ρ in the usual manner. All the calculations of the preceding section apply, except that since the position of standing wave minimum cannot be plotted as a function of wavelength, a phase curve similar to Fig. 5 cannot be obtained. Thus it remains to be determined whether the cavity is undercoupled or overcoupled. This coupling condition can be determined by a knowledge of loop orientation and the variation of ρ_0 at resonance as the coupling is altered. If ρ_0 decreases as the coupling is increased, the cavity is undercoupled; if ρ_0 increases as the coupling is increased, the cavity is overcoupled. For the undercoupled case, $g_0=R_{V0}>1$, and for the overcoupled case $g_0=1/R_{V0}<1$.

It often happens that the frequency cannot be perturbed sufficiently to make the required excursions. In that case, the Q can be determined by measurements near resonance in the following manner.

Since R_∞ cannot be measured now, it is necessary to make some assumptions about the series loss g_s. In most

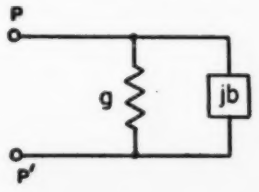


FIG. 10. Equivalent circuit for $g_s \to \infty$.

[5] J. F. Morrison and E. L. Younker, Proc. Inst. Radio Engrs. 36, 212 (1948).

cases at these frequencies, it is justifiable to neglect it, as is done here. Thus Q_U is given by Eq. (29), and since $g=g_0$ is determined at resonance, it only remains to measure β. Slightly off resonance, the cavity admittance is

$$y=g+jb=g+2j\beta\Delta\omega/\omega_0. \tag{31}$$

Here it is assumed that the frequency excursion $\Delta\omega \ll \omega_0$ so that $\omega/\omega_0-\omega_0/\omega$ is equal to $2\Delta\omega/\omega_0$ very closely. The modulus of the reflection coefficient ρ is given by Eqs. (6) and (9) as

$$\rho^2=[(1-g)^2+b^2]/[(1+g)^2+b^2], \tag{32}$$

where

$$\begin{aligned} b^2&=[\rho^2(1+g)^2-(1-g)^2]/(1-\rho^2)\\ &=\left[R_V{}^2-R_V\left(g+\frac{1}{g}\right)+1\right]g\Big/R_V\\ &=P_r(1+g)^2-P_i(1-g)^2/(P_i-P_r). \end{aligned} \tag{33}$$

Thus b may be calculated using the most convenient form of Eq. (33). Then β is calculated by reference to Eq. (31) and Q_U by Eq. (29); for example,

$$Q_U=\omega_0/2g\Delta\omega[P_r(1+g)^2-P_i(1-g)^2/P_i-P_r]^{\frac{1}{2}}. \tag{34}$$

P_i and P_r are to be measured at the frequency excursion corresponding to $\Delta\omega$. If the cavity is critically coupled at resonance, $g=1$ and Eq. (34) becomes

$$Q_U=(\omega_0/\Delta\omega)[P_r/(P_i-P_r)]^{\frac{1}{2}}. \tag{34a}$$

The solution for b may be carried out graphically using a normalized admittance chart containing circles of constant reflection coefficient. As seen from the above discussion, there are two possible values of g, and the proper one is chosen by a knowledge of the condition of the coupling. Graphically, it can be shown that the two values of g are the two intercepts on the g axis of the circle of constant reflection coefficient ρ_0. At a frequency slightly different from resonance the new value of $\rho=(P_r/P_i)^{\frac{1}{2}}$ determines a second circle on the chart. The intersection of this circle with the appropriate line $g=$constant determines b.

In practice, several readings are made at different frequency excursions and corresponding values of P_i and P_r. The various calculated values of Q_U are then averaged. It has been assumed that g_s was infinite; whether this is indeed the case depends on construction of the coupling device. If frequency and power output of the signal source are stable and the cavity is reasonably rigid, good results are obtained within 5 percent accuracy. Care must be exercised in choosing the frequency excursion so that P_r or P_i-P_r can be measured accurately. A wise choice can be made if the Q is known approximately. Measurements of P_r and P_i can generally be easily made, but measurements of the frequency excursion may be troublesome, for if the Q_U is of the order of 1000, variations of one part in 10^5 must be measured.

IV. MEASUREMENT OF A STEADY-STATE DISCHARGE

The introduction of a steady-state discharge into the cavity is equivalent to the addition of an admittance g_d+jb_d in parallel with the cavity elements $g+jb$, that is, in parallel with the parallel elements of Fig. 1(d). It is convenient to determine all the quantities except g_d and jb_d on the empty cavity prior to the initiation of the discharge. Theoretically two measurements will then suffice to determine these quantities and a complete resonance curve of the cavity containing the plasma need not be measured. The discharge tends to undercouple the cavity, to lower its Q, and to change its resonant frequency. Thus the db standing wave ratio R_0 at resonance ω_0 can be considered as changed to R_0' at ω_0' and these new quantities are measured. Since β and g_s do not change, measurement of R_0' yields the unloaded Q of the cavity with the discharge going as

$$Q_U'=\beta(Q_L'/\beta)(Q_U'/Q_L'). \tag{35}$$

From these plotted ratios, using R_0' and the known β and R_∞, Q_U' can be determined.

Such measurements involve frequency excursions of the applied field of the order of one percent. This has a negligible effect on the discharge except in so far as the amount of power transmitted to the discharge by the cavity is concerned. It is therefore useful to monitor the light output of the discharge so that the input power can be adjusted for constant light intensity and hence constant discharge density as the frequency is varied.

The authors wish to acknowledge the valuable assistance of Dr. Melvin A. Herlin and Dr. Edgar Everhart in the preparation of M.I.T. Research Laboratory of Electronics Technical Reports 66 and 140. The techniques presented in this paper were taken in part from these two previous reports.

Methods of Measuring the Properties of Ionized Gases at High Frequencies. II. Measurement of Electric Field*

DAVID J. ROSE† AND SANBORN C. BROWN
Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received January 24, 1952)

Experimental methods are given for determining the electric field within resonant cavities in the microwave region. Methods are discussed of calculating the electric field in both simple and complex structures used in gas discharge studies.

I. INTRODUCTION

ONE of the most important physical parameters which control the behavior of a discharge of electricity in gases is the electric field in which the discharge takes place. The methods of determining the electric field at high frequencies are therefore of great importance in the study of the properties of ionized gases at these frequencies.

The electric field E is determined in terms of the incident power on the cavity P_i and the standing wave pattern on the input line. Three general procedures will be described. The first two are applicable to high Q_L cavities. The first of these is applied to cavities in which the field configuration can be computed easily while the second is applicable when the exact configuration is not known. The third method applies to low Q_L cavities. In some cases it is desirable to record the incident power at frequencies other than the resonant frequency of the cavity, and the conversion of such measurements to equivalent ones at resonance is discussed. The notation of Part I of this series of papers is used.

II. THE SIMPLE MODE METHOD

Consider the magnitude E_0 of the rms electric field at some chosen reference point in the empty cavity. The stored energy can be expressed as

$$U=\eta E_0^2, \tag{1}$$

where η is a parameter fixed for a given mode of oscillation and reference point. From the relation $U=CV^2$, where C is the capacitance of a lumped-constant resonant circuit and V is the rms voltage across it, it is seen that the parameter η is a capacitance-like quantity in which the energy is referred to the electric field instead of the voltage.

The determination of Q for resonant cavities was discussed in Part I of this series of papers.[1] If we use the definition of Q_U discussed there and the definition of η, we may write

$$E_0=(P_0Q_U/\eta\omega_0)^{\frac{1}{2}}, \tag{2}$$

where P_0 is the power absorbed at resonance. The quantity η must be computed for the cavity in question. This calculation is illustrated below for the cylindrical TM_{010}-mode cavity, where the reference point chosen is at the center. When the cavity has a complicated shape, it is not feasible to compute η, and the second method must be used.

For the cylindrical TM_{010} mode

$$E_r=E_\phi=0,$$
$$E_z=\sqrt{2}E_0J_0(2.4r/r_0)\exp j\omega t,$$

where r_0 is the radius of the cavity and E_0 is the rms value of the electric field at the center. At the instant the electric field is a maximum, the magnetic field (not given above) is zero and the stored energy is

$$U=0.5\epsilon_0\int_0^{r_0}[\sqrt{2}E_0J_0(2.4r/r_0)]^2 2\pi rh dr$$
$$=0.2695\epsilon_0\pi r_0^2hE_0^2,$$

where h is the height of the cavity and ϵ_0 is the permittivity of free space. Thus, $\eta=0.2695\pi\epsilon_0r_0^2h$.

III. TUNING RATE METHOD

For a cavity of all but the simplest shape, it is often difficult and sometimes impossible to compute η. For these cases, an experimental method is available for measuring the electric field directly at any point on the cavity surface where the magnetic field is zero. At such points, the electric field is a maximum.

Suppose a small metallic body of variable volume is inserted into the cavity in the region of maximum electric field; this volume might well take the form of a plug inserted through a hole in one of the surfaces. The insertion of a volume dv into the cavity will result in a shift $d\omega$ of the resonant frequency ω_0 given by Slater:[2]

$$d\omega=-\omega_0E_0^2dv\Big/2\int E_1^2dv, \tag{3}$$

where E_0 is the rms electric field at the point in question and E_1 is the rms electric field at any point in the cavity. One recognizes that $\int E_1^2dv=U/\epsilon_0$, and using this relation and the definition of Q as[1] $Q=\omega_0U/P$ where P is the

* This work has been supported in part by the Signal Corps, the Air Materiel Command, and the ONR.

† Now at Bell Telephone Laboratories, Inc., Murray Hill, New Jersey.

[1] S. C. Brown and D. J. Rose, J. Appl. Phys. 23, 711 (1952).

[2] J. C. Slater, Rev. Mod. Phys. 18, 482 (1946). The relation is obtained by noting that in this reference Slater's cavity field function E_a is normalized so that $\int_{\text{vol}} E_a^2dv=1$.

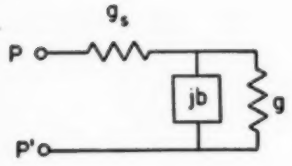


FIG. 1. Equivalent circuit of conductive and susceptive elements.

energy dissipated per second, Eq. (3) may be put in the form

$$E_0^2=(d\lambda/dv)Q_U P_0(\mu_0/\epsilon_0)^{\frac{1}{2}}/\pi, \tag{4}$$

where $(\mu_0/\epsilon_0)^{\frac{1}{2}}=377$ ohms. This relation is equivalent to defining η as

$$\eta=(dv/d\lambda)\pi(\epsilon_0/\mu_0)^{\frac{1}{2}}/\omega_0.$$

The quantity $d\lambda/dv$ may easily be measured, and it is to be evaluated at the point where the plug is flush with the surface through which it is inserted.

IV. DETERMINATION OF ABSORBED POWER AT RESONANCE

Since the formula for the electric field in Eq. (2) is given in terms of the power absorbed at resonance, it is necessary to convert the power incident on the cavity at any frequency P_i, which is the quantity measured, to the power absorbed at resonance P_0. Consider the circuit of Fig. 1 which is drawn in this simple form considering g and $jb=j\beta(\omega/\omega_0-\omega_0/\omega)$ to contain all the conductive and susceptive elements. Let P represent the power absorbed at any frequency. These elements may represent the empty cavity alone or may represent the empty cavity plus a discharge admittance g_d+jb_d. The only restriction is that if the cavity contains a discharge, the Q of the system must remain high, and the perturbation of the empty resonance ω_0 must be less than one percent. If the discharge is to remain constant with frequency, the power absorbed in g must be constant. For a constant power in g, it can be shown that

$$P_0=P/[1+b^2/gg_s(1+g/g_s)]. \tag{5}$$

In Part I we showed that $P=P_i(1-\rho^2)$ where ρ is the magnitude of the reflection coefficient at the frequency of measurement corresponding to b and further that the conductance across PP' at resonance $g_0=g/(1+g/g_s)$, so that we may express Eq. (5) as

$$P_0=P_i(1-\rho^2)/[1+b^2(1-g_0/g_s)^2/g_0g_s]. \tag{6}$$

The quantity b can be related to ρ, g_0, and g_s in the following manner. The reflection coefficient Γ at PP' of the circuit of Fig. 1 is

$$\Gamma=\frac{1+g/g_s-g-jb(1-1/g_s)}{1+g/g_s+g+jb(1+1/g_s)}. \tag{7}$$

This equation may be solved for ρ in terms of b. Substitution in Eq. (6) gives

$$P_0=ABP_i, \tag{8}$$

where

$$1/A=1+(1/g_s^2)-(1/g_s)(g_0+1/g_0) \tag{9}$$

and

$$B=(1-1/g_s)^2-\rho^2(1+1/g_s)^2. \tag{10}$$

The factor A is constant for any particular resonance curve and is identical for overcoupled and undercoupled cavities. In Part I we saw that g_0 and g_s could be expressed in terms of the resonant db standing wave ratio R_0 and the off-resonance ratio R_∞, respectively. In Fig. 2 the quantity A is plotted as a function of R_∞ and R_0. The factor B is a function of R_∞ and the actual standing wave ratio R_m (as determined by ρ^2) at which the particular measurement of P_i has been made. This is plotted in Fig. 3.

In practice, it is often found that the inductance associated with the coupling loop of the cavity causes the measured resonance to differ from the actual resonance of the cavity. This effect is recognized by the fact that the power P_0 calculated from Eq. (8) is not constant with frequency. It is sufficiently accurate to correct for this effect by displacing the resonance curve, but not the power measurements, slightly in frequency and calculating P_0 with the adjusted data. The amount of shift in a practical case is small, and because of the complicated nature of the equation if such a reactive element is inserted, it is best determined by trial as outlined above.

If P_i is measured at the resonant frequency ω_0, $P=P_i(1-\rho^2)$ yields P_0 directly. The top curve of Fig. 3 represents P_0/P_i as a function of R_0 in this case.

V. FOUR-TERMINAL NETWORK METHOD

In cases where the Q of the cavity is very low, it is not accurate to apply cavity theory because of the complications introduced by the modes of the cavity. Actually the discharge need not be located in a cavity at all, and may, for example, be placed in a wave guide. In such cases an accurate method exists for measuring the electric field, subject to the restriction that the region of the discharge is small compared to a wavelength.

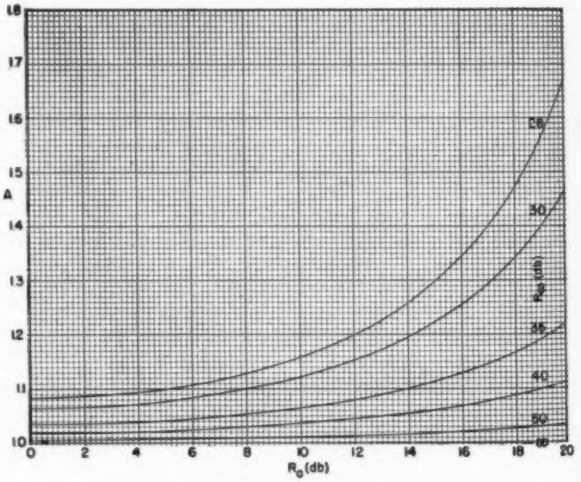


FIG. 2. Factor A plotted as a function of R_0 and R_∞.

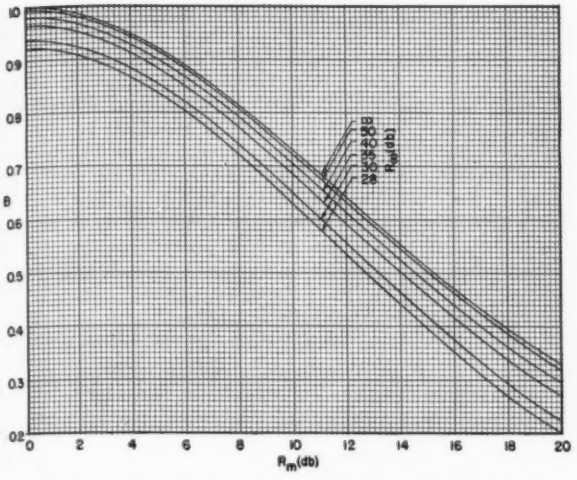


FIG. 3. Factor B plotted as a function of R_m and R_∞.

Let PP' be terminals located at some arbitrary point on the transmission line, and TT' be the terminals across which the discharge will take place. It is permissible to postulate these terminals since the transmission line supports only one mode of propagation and the discharge region is small compared to a wavelength. The equivalent circuit of Fig. 4 represents the circuit without a discharge. Here, the circuit elements are not normalized to the characteristic admittance Y_c of the line. The series loss G_s and the shunt loss G include the transmission line and cavity losses. The reactive elements are represented by a series reactance X_s, a shunt reactance B, and an ideal transformer of admittance ratio a^2:1. Different elements could be used but any lossless three-element set can be represented as shown. The specific introduction of the transformer is necessary, as an admittance transformation occurs between the transmission line and the discharge position. A similar transformation exists for high Q cavities, but the admittance ratio is incorporated into the measurements of the Q and η. Here, however, such calculations cannot be made.

The circuit parameters of Fig. 4 are determined experimentally by short-circuiting TT' and measuring the standing wave pattern on the line as a function of frequency, and by opening TT' and repeating the measurements. The position PP' is chosen as the position of the standing wave minimum with TT' short-circuited, at the frequency where the minimum with TT' open is exactly $\lambda_g/4$ away. This procedure also

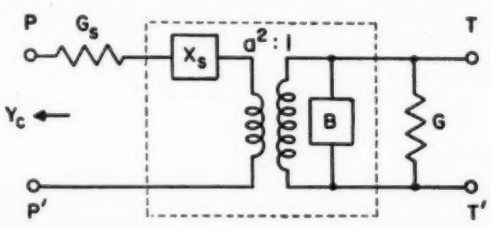


FIG. 4. Four terminal network representing a cavity.

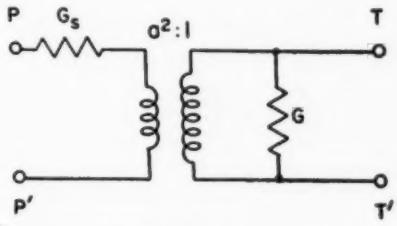


FIG. 5. Circuit of Fig. 4 simplified by correct choice of reference plane and frequency.

defines an angular frequency ω_0 characteristic of the empty cavity and line. The fact that PP' is at a minimum when TT' is short-circuited means that $X_s=0$, and the fact that PP' is $\lambda_g/4$ away from a minimum when TT' is open means that $B=0$ also. Thus at $\omega=\omega_0$ and this choice of PP', the circuit may be simplified to that of Fig. 5. The voltage standing wave ratio $R_{V\infty}$ with TT' short-circuited is $R_{V\infty}=G_s/Y_c$, and the ratio R_{V0} with TT' open is given by $R_{V0}=Y_c(1/G_s+1/a^2G)$.

A separate determination of a^2 is necessary and this is carried out by altering the capacitance across the terminals TT' and noting the change at PP'. The terminals TT' are the actual place where the discharge takes place and for ease in computation they are assumed to be parallel plates. The capacitance can be altered by adding dielectric or metallic disks, or by means of a movable metal plug set in one of the plates. The depth of such a plug must be much less than its lateral dimensions. This is similar to the measurement of electric field in a high Q cavity by the tuning rate method.

In its normal position, the plug is flush with the plate surface. If a small volume dv is inserted, the additional capacitance across TT' is $\epsilon_0 dv/h^2$ where h is the gap spacing. At $\omega=\omega_0$, all the reactive elements of the circuit cancel at PP' except for this small added capacitance. The normalized circuit at PP' takes the form of Fig. 6 where

$$g=a^2G/Y_c, \quad b'=a^2\omega_0\epsilon_0 dv/Y_c h^2. \tag{11}$$

In most cases the effect of g_s can be ignored, and we will do so here, though it may easily be included if necessary. In that case, the impedance $z_{PP'}\approx 1/(g+jb')$; and near resonance where $b'\ll g$, from Eq. (9), Part I, the phase angle ϕ of the reflection coefficient at PP' is

$$\phi\approx -2b'/(1-g^2). \tag{12}$$

This shift in ϕ results in a small change dx in the position of the minimum at PP', and the ratio dx/dv may be

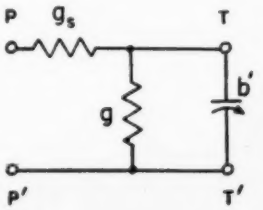


FIG. 6. Normalized equivalent circuit for the addition of a small capacity at TT'.

measured. Since

$$dx/dv=(dx/d\phi)\cdot(d\phi/db')\cdot(db'/dv)$$
$$=(\lambda_g/4\pi)[-2/(1-g^2)](a^2\omega_0\epsilon_0/Y_c h^2), \quad (13)$$

then

$$a^2=(\mu_0/\epsilon_0)^{\frac{1}{2}}(\lambda_0/\lambda_{g0})\,|dx/dv|\,(1-g^2)h^2Y_c. \quad (14)$$

The quantity $|dx/dv|$ must be evaluated at the point where the plug is flush.

This value of a^2 is applicable insofar as measurements are made at the points PP' at the frequency ω_0; that is, if the representation of Fig. 5 is to be maintained. In that case, the absorbed power $P_0=V^2G$ (neglecting G_s), where V is the actual rms voltage across the gap TT'. Thus, $V^2=P_0a^2/Y_cg$. But since PP' is at a standing wave maximum with TT' open, $R_{V0}=1/g$. The voltage across the gap is then given by

$$V^2=P_0a^2R_{V0}/Y_c; \quad (15)$$

for our parallel plate gap, the field $E=V/h$ is

$$E^2=(\mu_0/\epsilon_0)^{\frac{1}{2}}(\lambda_0/\lambda_{g0})\,|dx/dv|\,(1-g^2)P_0R_{V0}. \quad (16)$$

In case the series conductance must be taken into account, Eq. (16) becomes

$$E^2=\left(\frac{\mu_0}{\epsilon_0}\right)^{\frac{1}{2}}\frac{\lambda_0}{\lambda_{g0}}\left|\frac{dx}{dv}\right|\left[1-\left(\frac{gg_s}{g+g_s}\right)^2\right]\left(R_{V0}-\frac{1}{g_s}\right)P_0. \quad (17)$$

All the necessary circuit parameters have been determined. Impedance measurements at PP' in Fig. 5 will then give the value of any impedance added across TT'.

The authors wish to acknowledge the valuable assistance of Dr. Melvin A. Herlin and Dr. Edgar Everhart in the preparation of Massachusetts Institute of Technology Research Laboratory of Electronics Technical Reports 66 and 140. The techniques presented in this paper were taken in part from these two previous reports.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

The Initial Conduction Interval in High Speed Thyratrons

JAMES B. WOODFORD, JR., AND EVERARD M. WILLIAMS
Department of Electrical Engineering, Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received February 14, 1952)

The initial conduction interval in thyratrons, or time interval during which the major rise in current occurs, is shown to be dependent on two parameters. The first of these is dependent on tube construction and the second on circuit elements. It is shown that both factors are significant in operation of hydrogen thyratrons in millimicrosecond switching service.

THE ionization (and de-ionization) times[1,2] of the cathode-anode conduction path in thyratrons have received considerable study. De-ionization time has been given major emphasis[3] for many years because of the limitations it imposes on anode supply impedance and frequency in control circuits and its effect on repetition rate in pulse circuits. In comparison with de-ionization time, ionization times are so short as to have been regarded as negligible. Recent studies of hydrogen thyratron millimicrosecond pulse sources have shown, however, that inherent tube ionization time imposes distinct limitations on such generators and an accurate knowledge of ionization time and its control is desirable.

The problem of measuring or calculating the ionization time of thyratrons has been given some attention in the literature.[3-7] Calculations which agree qualitatively with experimental results have been made by Mullin[5] on the basis of a simplified thyratron model in which conduction results from ionization of gas molecules in the vicinity of the anode by electrons accelerated by the anode-cathode potential. In this model the primary cause of the limited rate of rise of anode current is the finite time required for electrons to move from the control grid to the anode and positive ions to move from the anode to the cathode region in order to neutralize the negative space charge surrounding the cathode. Mullin divides ionization time into two sub-intervals; during the first of these, here termed the delay period, the anode current changes very little, while during the second, here termed the commutation interval, most of the change in anode current occurs. The delay period was observed to depend roughly on the grid over-voltage, i.e., the voltage applied to the grid in excess of that required to trigger the thyratron, while the duration and behavior of the commutation interval were found to be closely independent of grid

[1] Wheatcroft, Smith, Metcalf, Phil. Mag. **25**, 649 (1938).
[2] E. T. Wheatcroft and T. G. Hammerton, Phil. Mag. **26**, 684 (1938).
[3] M. Bernbaum, Am. Inst. Elec. Engrs. Trans. **67**, 209 (1949).
[4] A. H. Harris, Am. Inst. Elec. Engrs. Trans. **59**, 747–750 (1940).
[5] C. J. Mullin, Phys. Rev. **70**, 401 (1946).
[6] H. Klemperer, Archiv. Electrotechnik **27**, 322–328 (1933).
[7] L. B. Snoddy, Physics **4**, 366–371 (1933).

over-voltage. Mullin also concluded that the commutation time is diminished by increasing either anode-supply voltage or gas pressure, or by decreasing tube dimensions. This information supplies qualitative data concerning the rise of anode current in modern high speed thyratrons, but is not sufficient for the determination of the commutation time in a particular system. Furthermore, in common hydrogen pulse thyratrons, because of the high grid and anode voltages, there is some doubt as to the validity of Mullin's model. For instance, it has been observed experimentally that part of the grid-cathode region is ionized when the high (several hundred volts positive) grid triggering voltage is applied and that only under suitable conditions (sufficient duration of grid voltage pulse and sufficiently high anode voltage) does this ionization spread to include the entire anode-cathode region. A new physical model which will allow quantitative determination of commutation time has not been completely developed, although work is now proceeding in this direction.

This paper describes a strictly experimental study of the commutation interval in high speed thyratrons and supplies some analytical aids for predicting the behavior of thyratrons operating with a resistive or resistive and inductive load. Figure 1 shows a typical switching circuit and an equivalent circuit including significant circuit parameters. The commutation time will be specified in the following way. During the commutation period, the anode voltage falls from approximately full supply voltage E_b to nearly zero with reference to the cathode. The initiation and conclusion of commutation are difficult to locate, since conduction starts very slowly and the final current rate-of-change approaches zero gradually. It is convenient, therefore, to consider a specific time interval Δt_{ab} or the time required for anode voltage to drop from $(E_b-a|\Delta V|)$ to $(E_b-b|\Delta V|)$, where ΔV is the total change in anode voltage and for practical

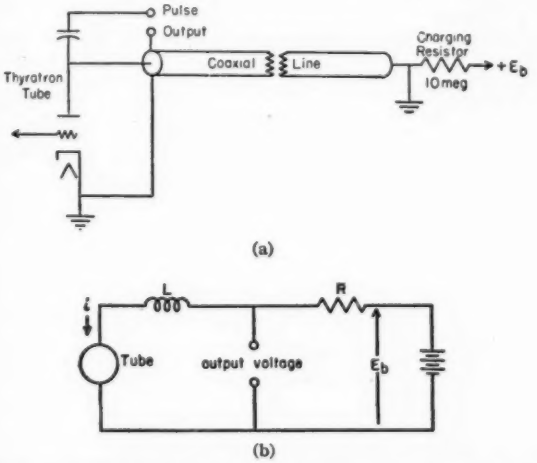


FIG. 1. (a) Typical thyratron switching circuit using a length of transmission line as a resistive load. (b) Electrical equivalent circuit during commutation interval.

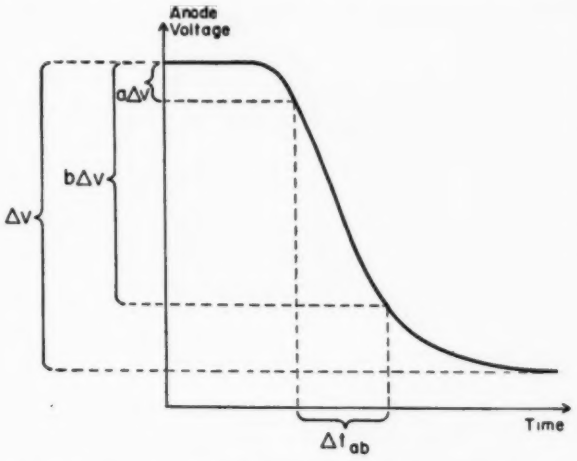


FIG. 2. Illustration of the definition of commutation time as employed in this paper.

loads and large thyratrons is approximately equal to plate supply voltage E_b. In all cases, $0<a<b<1$. If, by superposition, only the changing component of anode voltage V' is considered and initially V' equals zero, then Δt_{ab} is the time for V' to go from $a\Delta V$ to $b\Delta V$. Convenient values, that is, readily measured values for a and b, have been found experimentally to be respectively 0.1 and 0.8 or 0.9. This definition is illustrated in Fig. 2.

Present day ionization theory[8] does not supply reliable numerical means for calculating commutation time directly from tube and circuit parameters. The precedure followed was to fit a semi-empirical formula to the results of extensive experimental measurements of hydrogen thyratron commutation time under a variety of operating conditions. An expression of the form

$$i=i_0\epsilon^{(\alpha\int_0^t V\,dt)} \quad \text{or} \quad (1/i)(di/dt)=V, \tag{1}$$

where V is anode voltage, i is anode current, and i_0 and α are constants, was found to give good results. This expression differs from the results of an approximate theoretical derivation in that anode voltage appears to the first power instead of to the one-half power.

A differential equation can be derived, using (1), for the circuit of Fig. 1 with the solution

$$t_{ab}=(1/\tau)\log_\epsilon[b(1-a)/a(1-b)]+\sigma\log_\epsilon(1-a)/(1-b),$$

where

$$\tau=\alpha E_b \tag{2}$$

and

$$\sigma=L/R.$$

For the typical values of a and b of 0.1 and 0.9, respectively, (2) becomes

$$T_{ab}=(4.4/\tau)+2.2\sigma. \tag{3}$$

[8] See, for instance, L. B. Loeb, *Fundamental Processes of Electrical Discharge in Gases* (John Wiley & Sons, Inc., New York, 1939).

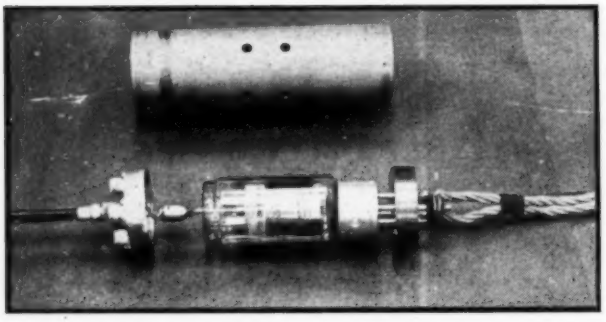


FIG. 3. Disassembled view of thyratron enclosure used to minimize lead inductance.

Thus the commutation time is the sum of two components—one dependent upon the reciprocal of the product of the anode supply voltage E_b and a parameter α and the other on the anode-circuit and load time constant.

Values of commutation time for a type 5C22 hydrogen thyratron were measured under a variety of operating conditions. In order to reduce the anode circuit time constant so that the effect of the tube ionization parameter α would be appreciable, the coaxial cavity-like enclosure (which minimizes lead inductance) shown in Fig. 3 was employed. Measurements were made photographically using a high voltage, high speed single-trace oscillograph. A typical trace appears in Fig. 4 and over-all results are shown in Figs. 5 and 6. Experimental curves are compared with curves plotted from Eq. (2), using the values of α which give the best agreement with all the experimental points. It will be noted that even at the relatively low anode voltages used, commutation times of the order of millimicroseconds are attainable. Furthermore, since the two parts of Eq. (2) are of the same order of magnitude in these thyratrons, it is evident that further increases in speed will require not only higher anode voltages, but also tube structure refinements to minimize lead inductance similar to those used elsewhere in the radio art.

FIG. 4. Typical thyratron commutation—anode voltage *versus* time. The time interval between marks is 65 milli-microseconds.

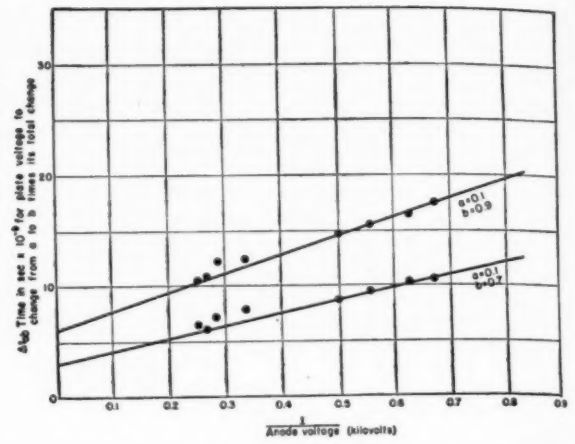


FIG. 5. Experimental determination of commutation time of a type 5C22 hydrogen thyratron with a constant load of 52 ohms and a minimum of series inductance. The solid lines were calculated from Eq. (2).

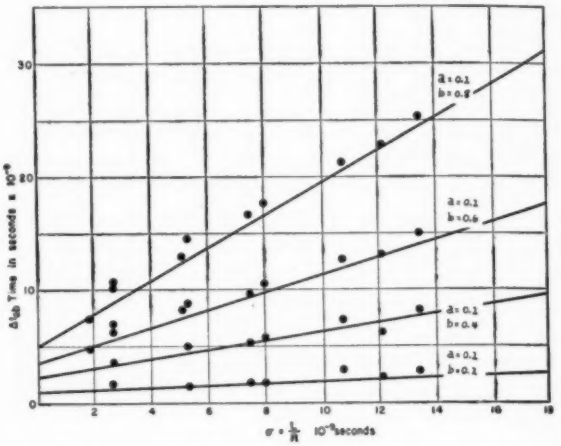


FIG. 6. Experimental determination of commutation time of a type 5C22 hydrogen thyratron with a constant anode-supply voltage of 3 kv and varying load resistance-series inductance time constant. The solid lines were calculated from Eq. (2).

The parameter α depends on tube geometry and pressure and type of gas or vapor employed. This parameter, at the present time, must be determined experimentally; it is known to vary from tube to tube of a particular type as well as from one tube type to another. Once determined experimentally, it may be used to predict the effect of changes in anode voltage and load time-constant.

On Diffusionless Transformation in Au-Cd Single Crystals Containing 47.5 Atomic Percent Cadmium: Characteristics of Single-Interface Transformation

LO-CHING CHANG
School of Mines, Department of Metallurgy, Columbia University, New York, New York
(Received January 24, 1952)

The kinetics of single-interface motion in the diffusionless transformation of Au-Cd single crystals containing 47.5 atomic percent Cd were studied. A diffusion-controlled relaxation phenomenon was found with an activation energy of the order of 24,000 calories per mole .The two observed characteristics of diffusionless transformation of this alloy, that the transformation does not take place isothermally, and that the velocity of transformation on cooling is considerably and consistently larger than that on heating, can be successfully interpreted from the relaxation hypothesis.

I. INTRODUCTION

THE characteristics of diffusionless transformation in Au-Cd single crystals containing 47.5 atomic percent Cd were reported previously.[1,2] The transformation takes place either by the formation of many domains of crystallographically equivalent orientations (multiple-interface transformation) or by the formation of a single domain of the low temperature phase (single-interface transformation). It is the purpose of this paper to describe some of the characteristics of single-interface transformation of this alloy.

II. THE VELOCITY OF SINGLE-INTERFACE MOTION

In the case of single-interface transformation, the transformation begins by the formation of a plate of the new phase at one end of the specimen, thereby creating an interface between the old and new phases, and continues by the relatively slow movement of this interface toward the other end of the specimen until complete transformation.

Motion pictures of the interface movement were taken. The rate of movement of the interface was studied by projecting the film, one frame at a time, on a screen provided with scales. Typical examples of the rate curves on heating and cooling are shown, respectively, in Figs. 1 and 2. The irregularity of interface motion is obvious. The jumping characteristics are more pronounced on heating than on cooling, especially at slow heating rates.

The average velocity* of the single-interface motion on heating or cooling was determined by measuring the time required by the interface to travel between two cross-hairs under a microscope stage while the specimen was being heated or cooled uniformly at a constant rate. Each measurement involved enough jumps so that the average values represent truly the slopes of straight lines running through the experimental points shown in Figs. 1 and 2. The temperature of the specimen was followed by a Brown Electronik Recorder. It was found that (1) for a given experiment the interface velocity is almost directly proportional to the rate of change of temperature and is considerably larger on cooling than on heating, and (2) for different experiments performed on the same specimen at various locations or on different specimens the interface velocity varies. The reason for this is not clear but probably is attributable to the state of internal conditions of the specimen (such as internal stresses, imperfections, composition fluctuations, etc.) and the environment of experimentation. The large variation in interface velocity for two separate experiments may be seen in Fig. 3.

III. RELAXATION EXPERIMENTS

It was thought that the jumping characteristics of interface motion is probably caused by some relaxation

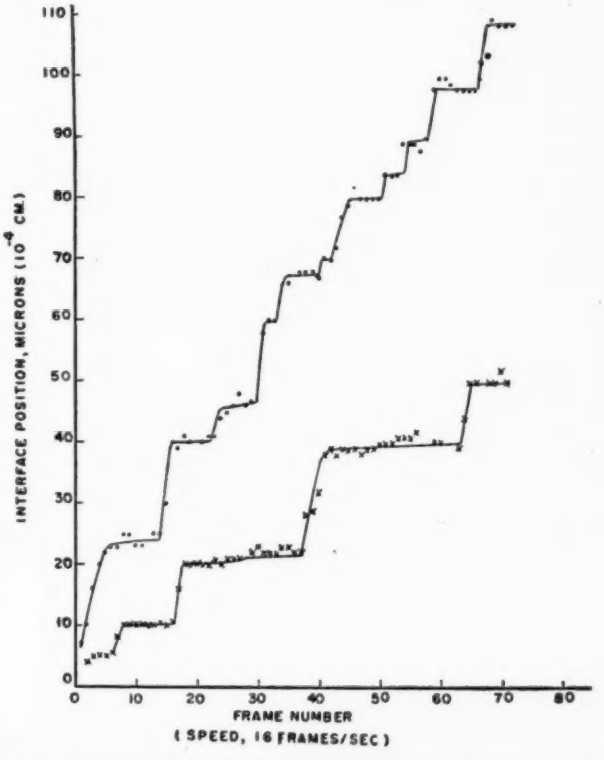


FIG. 1. Frame by frame analysis of interface motion on heating at two different heating rates.

[1] L. C. Chang and T. A. Read, J. Inst. Metals **191**, 47 (1951).
[2] L. C. Chang, Acta Cryst. **4**, 320 (1951).
* Corrected to velocity of propagation in a direction perpendicular to the interface.

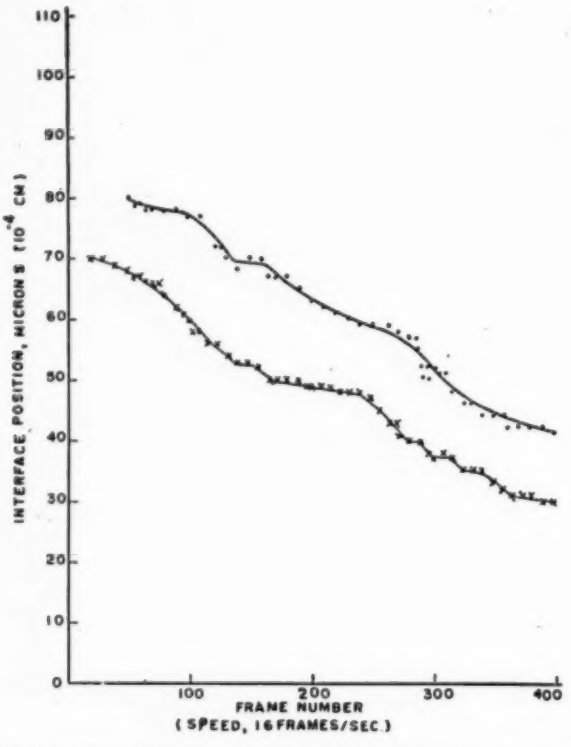


FIG. 2. Frame by frame analysis of interface motion on cooling from two independent runs.

process occurring at the interface between the two phases. Experiments were planned, therefore, to explore this relaxation phenomenon.[3] The experimental procedures used are illustrated in Fig. 4. The specimen under investigation was heated or cooled uniformly inside a relatively large copper block provided with openings for insertion of the specimen and microscopic observation of the interface motion. The temperature of the specimen was followed by a Brown Electronik Recorder. The interface on cooling was held at a certain position x at temperature T for a time interval t and released by resuming the cooling (see Fig. 4(a)). The specimen was then heated immediately at a constant heating rate; it was observed that the interface stops with a jerky motion at exactly the same position x for a time period t' and then proceeds with another jerky motion (see Fig. 4(b)). The previous holding on cooling presumably causes a redistribution of the potential barrier at the position x by some relaxation process, and the extra free energy, ΔF, required to surpass the increased potential barrier on reheating must come from an increased temperature differential, ΔT, which in turn is proportional to t' (see Fig. 4(b)). The amount of relaxation is therefore measured by t'. Since the exact relaxation mechanism is not known, a rigorous treatment of the problem is not yet feasible. One may, however, express the amount of relaxation by the following general function:

$$t' = f[te(-Q/RT)]. \tag{1}$$

For a given amount of relaxation, i.e., for a given t' (see Fig. 4) $te(-Q/RT)$ is constant. The data for t' and t at three temperatures are plotted in Fig. 5, showing that the ratio of t for any two temperatures at any given t' (between the limits of about 10 seconds to 35 seconds) is approximately constant. If for a given t' value the t values at two temperatures T_1, T_2 are, respectively, t_1, t_2, it follows:

$$\log \frac{t_1}{t_2} \cong \frac{Q}{2.303R}\left(\frac{1}{T_1} - \frac{1}{T_2}\right). \tag{2}$$

The Q value evaluated from Fig. 5 according to Eq. (2) is

$$Q \cong 24{,}000 \text{ calories/mole}.$$

Since Q is evaluated from one set of data obtained within a temperature range of about 6 degrees, the evaluated activation energy of 24,000 calories per mole serves only to indicate its order of magnitude. Further measurements will be necessary in order to obtain more accurate values.

Another demonstration of the relaxation phenomenon is seen in the following experiment. A single crystal of the Au-Cd alloy, about 4 cm in length, was suspended

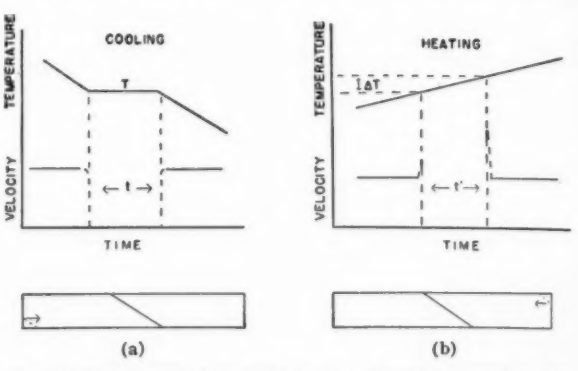


FIG. 4. Experimental procedures for relaxation experiments.

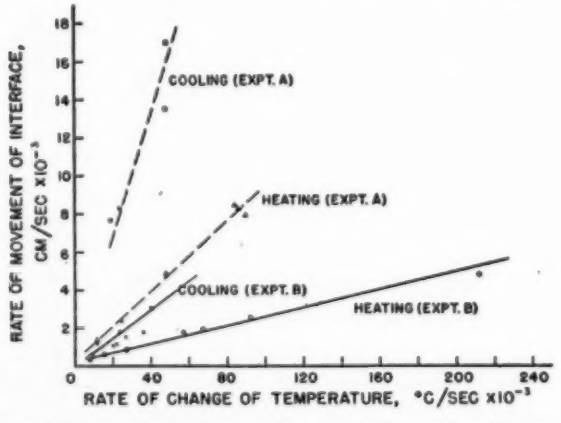

RATE OF MOVEMENT OF INTERFACE, CM/SEC X10⁻³
COOLING (EXPT. A)
HEATING (EXPT. A)
COOLING (EXPT. B)
HEATING (EXPT. B)
RATE OF CHANGE OF TEMPERATURE, °C/SEC X10⁻³

FIG. 3. Velocity of interface motion on heating and cooling *versus* rate of change of temperature.

[3] L. C. Chang and T. A. Read, Phys. Rev. 82, 770 (A) (1951).

vertically and transformed by heating from the ends. One thermocouple, A, was attached to the specimen at about 1 cm from the top end and a second thermocouple, B, was attached to the specimen at about 1 cm from the bottom end. Temperature readings of the thermocouples A and B were obtained when the transforming interface was reaching the respective positions where the thermocouples were attached. The specimen was first heated from the top end; two temperature readings $t_A{}^h$ and $t_B{}^h$, representing the transformation temperatures on heating at A and B, respectively, were then obtained. The heat was shut off and two more temperature readings, $t_B{}^c$ and $t_A{}^c$, representing the transformation temperatures on cooling at B and A, respectively, were obtained. Similar measurements were repeated while the specimen was heated from the bottom end instead of the top end. The results are shown in Table I.

If volume relaxation were absent, $t_A{}^h$ would be equal to $t_B{}^h$, and $t_A{}^c$ would be equal to $t_B{}^c$, at all times. The results given in Table I definitely show the existence of relaxation such that (1) the specimen does not transform isothermally and transforms only when the temperature

TABLE I. Comparison of transformation temperatures at A (1 cm from top end) and B (3 cm from top end) as the specimen was transformed by supplying heat to the top end and the bottom end.

	Heating		Cooling	
	$t_A{}^h$	$t_B{}^h$	$t_A{}^c$	$t_B{}^c$
Heat supplied to top end of specimen	57.1°C	58.5°C	48.4°C	49.1°C
	57.2	58.5	48.5	49.2
	57.0	58.4	48.2	49.2
Heat supplied to bottom end of specimen	58.5	57.1	49.0	48.1
	58.5	57.1	48.9	47.9
	58.5	...	48.9	47.7

of the specimen is continuously changing, and (2) the transformation temperature is not constant along the length of the specimen but varies according to the relative amounts of transformed and untransformed material.

IV. DISCUSSION

The above studies constitute experimental evidence of (1) a diffusion-controlled relaxation process taking place at the single-interface between the two phases during transformation, and (2) an additional volume relaxation process such that the resistance to transformation increases with the amount of prior transformation. The mechanism of these relaxation processes is not yet clear. Recent x-ray diffraction studies suggest that both processes may be related to relaxation of stress by reorientation of pairs of atoms. Results of these studies will be published separately. The effect of relaxation on the transformation behavior of the alloy is being studied in this laboratory.

The observation that for a given experiment performed on a given specimen the interface velocity on

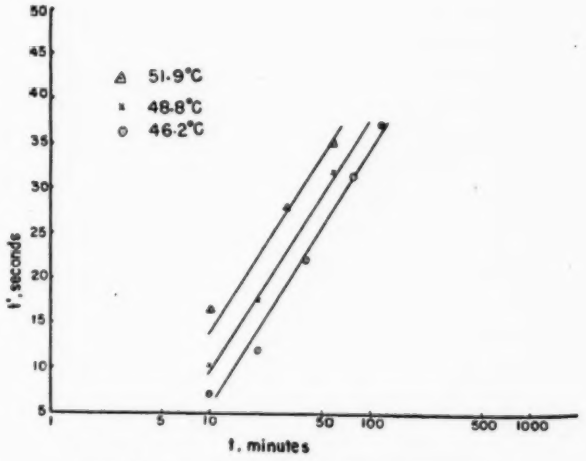


FIG. 5. Plot of t' *versus* t at three temperatures.

cooling is considerably larger than that on heating deserves some consideration. Since the relaxation process is strongly temperature dependent, one is tempted to think that this difference in interface velocity on heating and cooling may be interpreted as being a difference in the rate of relaxation arising from a difference in the transformation temperature. If one expresses the rate curves shown in Fig. 3 by the following empirical equation:

$$V=k(dT/dt), \tag{3}$$

where V is the observed interface velocity, dT/dt the observed rate of change of temperature, and k, the slope given by each straight line in Fig. 3, is a constant which is not only dependent on the transformation temperature but may also vary with the internal conditions of the specimen and the experimental environment. For a given experiment performed on a given specimen, k is probably only a function of the transformation temperature alone, and the effect of the rate of relaxation on interface velocity is probably the most important factor affecting k. If one assumes the following simple relation between k and the transformation temperature:

$$k\cong k_0 e(Q'/RT), \tag{4}$$

and further denotes heating by the subscript h and cooling by the subscript c, it follows:

$$\log\frac{k_c}{k_h}\cong\frac{Q'}{2.303R}\left(\frac{1}{T_c}-\frac{1}{T_h}\right). \tag{5}$$

Three sets of data from three independent experiments

TABLE II. Estimation of Q'.

Exp. No.	k_c, cm/°C	k_h, cm/°C	T_c, °K	T_h, °K	Q' cal/mole
A	.340	.095	330.6	341.0	27,000
B	.072	.024	333.4	343.1	25,000
C*	.156	.107	319.7	323.7	22,000

* Specimen containing about 48 atomic percent Cd.

are available to estimate Q' according to Eq. (5). The data and calculations are shown in Table II.

The estimated Q' values from three independent experiments agree fairly well with the Q value, 24,000 calories per mole, evaluated in Sec. III.

It may be mentioned here that if the specimen (see Fig. 4) on reheating is brought to and held at some other temperature Tx within the hysteresis range for a time interval tx and then heated up at a constant heating rate, the time of stoppage, t', of the interface at position x should decrease with increasing Tx and tx. Several experiments were performed. For given t and T values (interface holding on cooling), t' decreases with increasing tx and Tx values as predicted. The data are, however, not sufficiently accurate for an estimation of the activation energy for this process, although one set of calculations shows that it is of the order of 25,000 calories per mole.

ACKNOWLEDGMENT

Mr. Charles Braun made the interface measurements presented in Figs. 1 and 2. Mr. J. Intrater constructed apparatus and obtained the data presented in Table I. This work was supported by the Atomic Energy Commission under contract No. AT (30-1)-904.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Specimen Charging in the Electron Microscope and Some Observations on the Size of Polystyrene Latex Particles*

S. G. ELLIS
RCA Laboratories Division, Princeton, New Jersey
(Received February 27, 1952)

From observations in the shadow electron microscope and the standard electron microscope it is concluded that specimen charging does not significantly change the magnification of the latter, provided that the specimen is less than a few microns in thickness and the illuminating electron beam simultaneously strikes a nearby grounded conductor such as the supporting mesh.

The difference in size of shadowed and unshadowed Dow Latex 580G, lot 3584 reported by Kern and Kern (J. Appl. Phys. **21**, 705 (1950)) is attributed to an added layer of material on the shadowed particles. Evidence in support of this view is presented, and the results of other workers are discussed in the light of it.

INTRODUCTION

S. F. KERN and R. A. Kern[1] have shown that under certain conditions the images of shadowed polystyrene latex particles are larger than the images of unshadowed latex particles.† They suggest that this difference in apparent size may be due to electrostatic charging of the unshadowed particles. If this hypothesis were correct it would be a matter of concern to electron microscopists in general. There are, however, reasons to believe that a different explanation can be given for the observations reported by the Kerns.

In the first place experiments with a shadow electron microscope suggest that as normally operated the effects of specimen charging in the common form of electron microscope are negligible. Since these experiments are not easily repeatable in the ordinary electron microscope they are described in some detail below.

Secondly, Scott[2] and Cosslett[3] have reported values for the diameter of 580G latex which differ appreciably from the values obtained in most laboratories.[4] In particular, Scott finds that if a second drop of latex suspension is added to a previously prepared mount, two sizes of particles appear (see Fig. 1(d) of reference 2, in which the latex particles are *not shadowed*). This observation could be explained if material from the latex suspension was deposited on the latex particles in the final stages of drying so that the latex particles first mounted received two layers of deposit and those from the second drop applied received only one added layer. The Kern's experiment that is least susceptible to disturbance by bombardment-induced contamination involves adding latex suspension to a previously prepared and shadowed mount. Some experiments were, therefore, undertaken to test the possibility that material was being deposited on the latex particles during mounting.

OBSERVATIONS WITH THE SHADOW ELECTRON MICROSCOPE

In the shadow electron microscope Fig. 1(a) a lens L forms a reduced image of an electron source at P close to the specimen. A shadow image of the illuminated part of the specimen appears on a fluorescent screen

* This paper was presented, in part, at the meeting of the Electron Microscope Society of America held in Philadelphia, November 8 to 10, 1951.

[1] S. F. Kern and R. A. Kern, J. Appl. Phys. **21**, 705 (1950).

† The latex referred to here is Dow Latex 580G, lot 3584.

[2] G. D. Scott, J. Appl. Phys. **20**, 417 (1949).

[3] V. E. Cosslett, Proceedings of the Conference on Electron Microscopy, Delft 1949, p. 64 (published at Delft, 1950).

[4] C. H. Gerould, J. Appl. Phys. **21**, 183 (1950).

F. S. One might then expect the magnification to be given by the ratio: the distance from *F. S.* to *P* divided by the distance from the specimen to *P*. The area of the brightly illuminated region on the fluorescent screen depends on the area of the diaphragm *A* and is only slightly dependent on the focal length of the lens *L* (in the absence of a specimen) for the small range of focal lengths that give large magnifications.

If *P* is focused at the specimen [Fig. 1(b)], giving infinite magnification, and if the specimen is conducting the diameter of the illuminated area of the fluorescent screen is D_0, this diameter remains appreciably the same whether the lens is slightly strengthened [Fig. 1(a)] or is slightly weakened [Fig. 1(c)].

Workers with the shadow electron microscope have observed a different phenomenon when the specimen is mounted on an insultating film and is positioned so that no grounded conductor in its vicinity (such as the supporting mesh) is illuminated by electrons. With the lens strengthened [Fig. 1(a)] the diameter of the illuminated region of the fluorescent screen is less than D_0, and on weakening the lens this diameter increases abruptly at the point of infinite magnification [Fig. 1(b)] to a value greater than D_0 [Fig. 1(c)]. These observations are consistent with the production of a *positive* charge on the insulating film.

If the microscope is operating under the conditions of Figs. 1(a) or 1(c), with an insulating film as specimen, and if the specimen is moved so that a grounded supporting mesh enters the electron beam, the area of illumination at the fluorescent screen quickly assumes the diameter D_0. This indicates that the positive charge on the film has been neutralized..

The mechanism whereby the positive charge is produced is as follows: The primary electrons leaving *P* traverse the specimen with negligible probability of absorption, but during their passage through the specimen they eject secondary electrons from it leaving the illuminated part of the specimen with a net positive charge. This process continues until the potential of the illuminated part of the film is high enough to reduce the rate of loss of secondary electrons to a value which is balanced by the leakage of electrons into this region and by the capture of primary electrons.

It is evident that with a sufficiently thick film, a negative charge should be formed by the capture of primary electrons (the secondary emission ratio being less than unity at the voltages commonly used). An approximate value for the minimum thickness at which this will occur can be obtained from the Thomson-Widdington law[5]

$$\Phi_0^2-\Phi^2=(a\rho Z/A)x,$$

where

$e\Phi_0$ = energy of incident electrons,

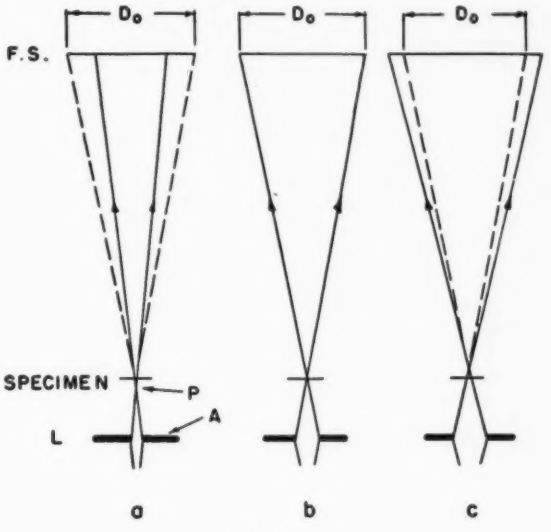


FIG. 1.

$e\Phi$ = average energy of electrons after passing through a layer of material of thickness x,

ρ = density of material,

Z = atomic number of material,

A = atomic weight of material,

and

a = a parameter that depends very little on atomic number and may be taken to be approximately 5×10^{11}.

To find the thickness of material for which the average exit velocity is zero we take $\rho=1.5$, $A/Z=2$, $\Phi_0=5\times10^4$ and $\Phi=0$, and obtain $x=67$ microns.

In other words, we may expect the charge on insulated specimens to be positive for all the thicknesses occurring in normal electron microscope practice, provided of course that there is no grounded conductor in the illuminating beam.

Several mechanisms may be advanced to explain the way in which the charge is neutralized when the mesh is illuminated. They are: (1) that the supporting film is rendered conducting by electron bombardment;[6] (2) that low velocity secondary electrons from the mesh ionize the residual gas in the microscope; (3) that low velocity secondary electrons from the mesh are attracted to the film and neutralize the positive charge directly. Hypothesis (1) is discounted by the observation that when using a fiber as specimen, charge neutralization can be obtained even when there is no solid connecting path between the illuminated part of the fiber and the illuminated part of the mesh. Hypothesis (2) is rendered unlikely by the paucity of negative ions produced by the electron bombardment of a gas. Hypothesis (3), therefore, remains as the most likely explanation of the phenomena observed.

It is difficult to make quantitative estimations from the observations reported above (the maximum deflec-

[5] See, for example, Zworykin, *et al.*, *Electron Optics and the Electron Microscope* (John Wiley and Sons, Inc., New York, 1945), p. 687.

[6] L. Pensak, Phys. Rev. **75**, 472 (1949).

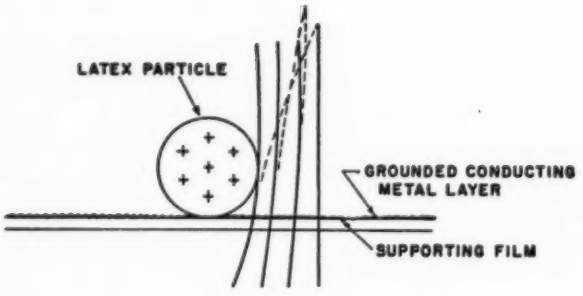


FIG. 2.

tions observed were of the order 10^{-3} radian). These observations do not prove that the positive charge is completely neutralized when the supporting grid is also illuminated. They do show that the charge is very considerably reduced. Qualitative considerations show that specimen charging would produce an apparent increase in the size of unshadowed latex particles. Consider a latex particle mounted on a previously prepared and shadowed mount of latex. (See Fig. 2.) If we suppose that the metal film on the substrate is conducting,‡ the electrons when below the film are not influenced by a charge on the latex particles. While above the film electrons would be attracted towards the latex particle. The force of attraction would decrease with distance from the latex particle and, hence, it would not have a lens action, which requires that the lateral impulse on the electrons increases in proportion to the radially measured distance of closest approach.[7] Instead there would be created a caustic surface, axially symmetric with respect to the center ray through the latex particle, so that if the objective lens were weakened a bright ring would appear outside the true edge of the latex particle. This phenomena has not been observed during the studies reported in this paper. It can sometimes be observed and is readily recognizable when working with insulating fibers in the electron microscope.§

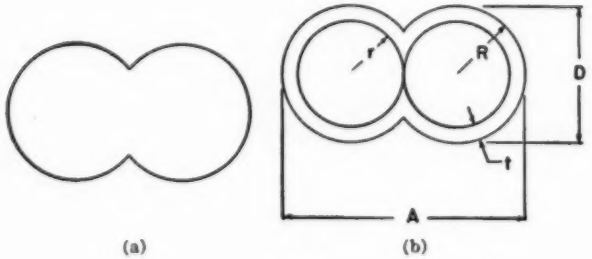


FIG. 3.

‡ Kern and Kern do not discuss the conductivity of their metal films. It seems possible that a platinum-palladium film less than 10A thick will be formed of discrete metal particles and will have a surface conductivity very little greater than that of the unshadowed substrate.

[7] K. Schlesinger, Proc. Inst. Radio Engrs. **32**, 483 (1944).

§ Since writing the above paragraph a letter has been published by J. H. L. Watson [J. Appl. Phys. **22**, 1387 (1951)] which illustrates this effect. Figures 1(f) and 1(g) of this letter show the electron concentration on the objective side of the specimen that would be expected from Fig. 2 above. It follows that the charge on the fiber is *positive*.

TABLE I. (All dimensions in A).

	D	d	t
Shadowed particles	2620	2340	140
Unshadowed particles	2520	2380	64

Since these observations do not support the charging hypothesis advanced by the Kerns, it is necessary to look for an alternative explanation for their observations.

OBSERVATIONS WITH LATEX PARTICLES

As stated in the introduction it was early suspected that some finely divided material was present, in addition to the latex particles, in the latex suspensions. When a drop of latex was dried on a collodion film, this additional material would coat the film and the latex particles making them look larger than they were in the original suspension. The evidence in support of this view comes from two sources: the appearance of particles in contact and direct electron microscope observation of added layers of material in the preparations.

Before proceeding with a description of the experimental work, it is necessary to interpose a word of caution. The results reported here are not precisely reproducible. In this they parallel the observations of Scott.[2] Indeed it has not been uncommon to find particles with markedly different appearances in different parts of the same mount, some particles being spherical and the others highly distorted. The measurements reported below were made on undistorted particles. Differences in apparent particle diameter may also occur between different mounts. This is believed to be due to differences in the ratio of volume to basal area in the drops added to the collodion film, and perhaps to differences in the rate of drying.

When two latex particles are observed in contact, they have the appearance shown in Fig. 3(a). This is not the appearance of two rigid spheres in contact since they do not touch at a point, tangentially. A possible explanation of this morphology is that two spheres of diameter $d(=2r)$ made point contact [Fig. 3(b)] and that later an added layer of thickness t was deposited on the particles. An isolated particle would then appear spherical with diameter $D(=2R)$.

TABLE II. (All dimensions in A).

Method of Mounting	D	d	t
From dilute suspension, unshadowed.	2580	2420	80
Using a centrifuge,* unshadowed.	2660	2470	95
Using a centrifuge, shadowed.	2650	2500	75
From acidified suspension, shadowed.	2640	2510	65
From alkaline suspension, shadowed.	2650	2430	110

* H. R. Crane, Rev. Sci. Instr. **15**, 253 (1944).

From Fig. 3(b)

$$d=A-D, \tag{1}$$

$$t=(D-d)/2. \tag{2}$$

This scheme of analysis was applied to the case in which a drop of latex suspension was applied to a previously dried and shadowed latex preparation with the results shown in Table I.

It is interesting to observe that the shadowed particles, which have twice been exposed to the liquid of the latex suspension, give a value of t approximately twice that of the unshadowed particles, which passed through only one drying operation.

In a further series of experiments single mounts of latex particles were made by different methods and the above method of analysis applied to double particles with the results shown in Table II. The average deviation in d in a given experiment was approximately ± 50A. In view of the variations in morphology that could sometimes be found within a given mount, no great significance is attached to the variations in D and d. The values of d are not, however, influenced by the presence of bombardment induced contamination.

The magnification of the electron microscope was obtained by photographing one mesh opening of a latex mount first in the electron microscope and then in the light microscope, the light microscope magnification having been determined by photographing a micrometer scale.

In contradistinction to the assumptions made above, Dillon, Matheson, and Bradford[8] have attributed the morphology of double latex particles to fusing or sintering. They analyze their results according to a relation due to Frenkel:

$$\theta^2=3\gamma t/2\eta, \tag{3}$$

where

t = time for which the particles are in contact,
γ = surface tension of the latex particles,
η = viscosity of the latex particles,

and

θ and R are defined in Fig. 4(a).

Since $\gamma t/\eta$ is constant in any latex mount, the following relation should hold for any one mount of a latex with different particle sizes:

$$\theta^2=\text{constant}/R. \tag{4}$$

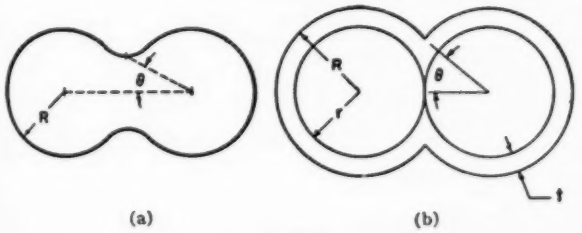


FIG. 4.

[8] Dillon, Matheson, and Bradford, J. Colloid. Sci. 6, 108 (1951).

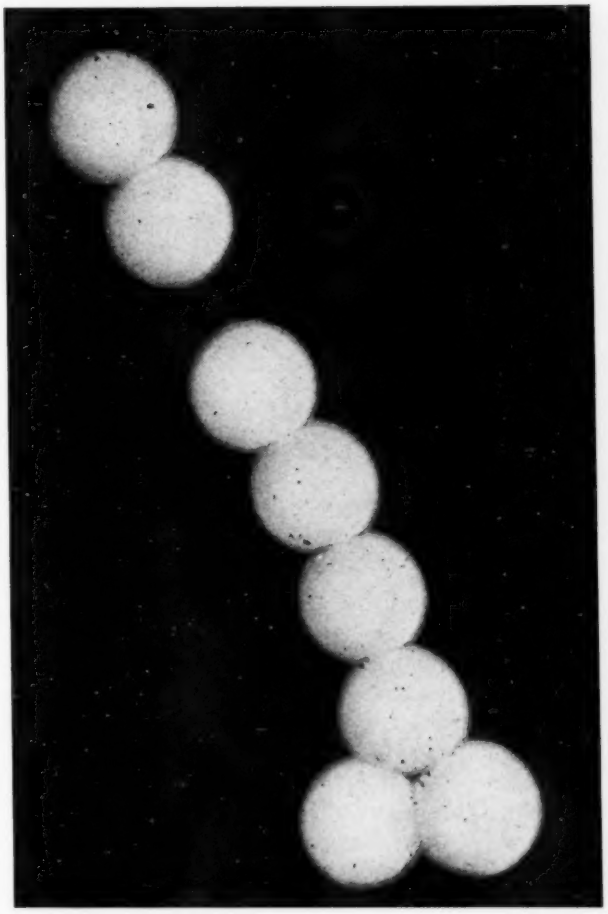


FIG. 5.

On the other hand if one applies the ideas illustrated in Fig. 3(b),

$$\theta^2=2[2R-t]\cdot t/[R-2t]\cdot R, \tag{5}$$

and if $t \ll R$,

$$\theta^2=4t/R. \tag{6}$$

The fact that Dillon *et al.* obtain a linear relation between θ^2 and $1/R$ does not therefore serve to distinguish between the two hypotheses. In fact, their results for polystyrene latex could be explained with $t=85$A. A similar ambiguity exists in considering the relation of particle radius to shadow length in the case of shadowed specimens.

Direct evidence for the existence of added layers can be found in some electron micrographs. Figure 5 is an example. Even here, however, there is a possibility of ambiguity. It is known that in some cases latex particles have diameters smaller than the maximum shadow width. If such a particle is covered with a layer of bombardment-induced contamination, its apparent diameter may nearly equal the maximum shadow width; yet the added layer is an artifact produced during observation in the electron microscope.

A more convincing experiment is to mount some

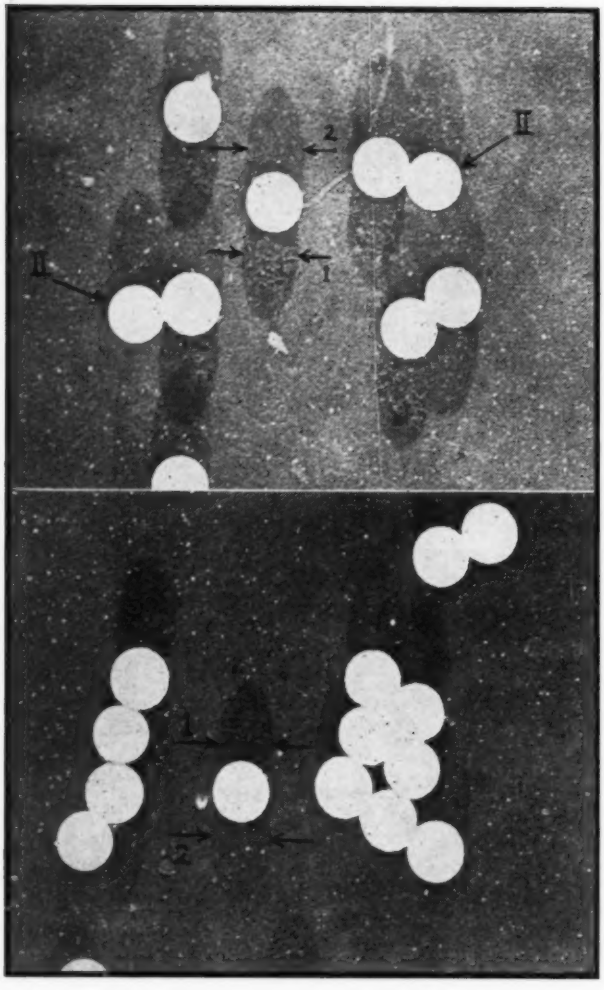


FIG. 6. II indicates latex particles in the second lot. 1 indicates the first, and 2 the second shadowing.

latex particles (I), shadow them, add a second lot of latex particles (II), and reshadow in the opposite direction. Careful measurement (Fig. 6) has disclosed that the second shadows (2) are slightly wider than the first (1). The second lot of particles are readily distinguishable from the first, since they have a smoother profile. Finally in the region of the substrate unprotected by the first shadow, it appears that surface material has been removed in small patches by the liquid of the second drop (Fig. 6). This effect, like the change in diameter, is not strictly reproducible.

The probable reason for the variation in the results is that the extra material in the drop of latex is deposited both on the collodion substrate and on the latex particles. The thickness of the layer formed will depend on the ratio of basal area to volume of the drop of suspension added to the film. It will also depend, in part, on the relative surface areas of the totality of latex particles, and substrate film presented to the drying drop of suspension, and may depend on the proximity of the field examined to the edge of the drop and the rate of drying. Finally, the general condition of the latex suspension will be important. In these experiments the added layer seemed to be thinner when using an old suspension in which there was evidence of bacterial contamination.

It is interesting to notice that the diameter as determined with the ultracentrifuge (2520±25A) is smaller than the majority of electron microscope results (range 2520 to 2630A, average 2588A).[4] This is in apparent agreement with the hypothesis advanced above. The value for $2r$ given below is less than the diameter found with the ultracentrifuge. However, one must be cautious in comparing the ultracentrifuge and electron microscope values for the diameter since the particles may shrink on drying.

The value obtained by small angle x-ray scattering[9] (2780A) is larger than most electron microscope values, but it is not larger than might be expected from the method of specimen preparation employed.

CONCLUSIONS

No evidence has been found to support the hypothesis that the apparent magnification in the electron microscope is influenced by specimen charging, provided that the specimen is less than a few microns in thickness and provided that the grounded, supporting, conducting mesh is illuminated with electrons.

Some of the variations in the apparent size of polystyrene latex particles, reported in the literature, are due to the deposition of an additional layer of material on to the latex spheres. This material was dispersed in the latex suspension and is probably deposited in the final stages of drying.

The diameter of the latex particles, prior to the addition of the extra material [the $2r$ of Fig. 3(b)] was found to be 2420±50A. It would be interesting to see if other observers, using this method of analysis, get a similar value.

[9] K. L. Yudowitch, J. Appl. Phys. **22**, 214 (1951).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Theory of Wave-Guide-Fed Slots Radiating into Parallel-Plate Regions

HARRY GRUENBERG
Radio and Electrical Engineering Division, National Research Council, Ottawa, Canada
(Received January 28, 1952)

Slotted wave-guide arrays feeding into parallel-plate regions have been used in some high speed scanners. Parallel-plate regions also have been used for the suppression of second-order beams of high gain arrays.

A theoretical expression is derived for the conductance of a longitudinal shunt-slot in a rectangular guide when the slot is radiating into a parallel-plate region of arbitrary plate spacing. Some peculiarities of the theoretical results are discussed. There is good agreement between theory and experiment.

INTRODUCTION

PARALLEL-PLATE regions in conjunction with slotted wave-guide arrays have been used in high gain antennas for the purpose of suppressing second-order beams[1] and also in some high speed scanners.[2] Hence a formula for the slot conductance under these conditions is of some importance.

According to Stevenson,[3] a longitudinal slot in the broad face of a rectangular guide is equivalent to a lumped shunt admittance across the guide at the center of the slot. If the slot is resonant, i.e., approximately half a free-space wavelength long, its normalized conductance is given by

$$g_0=2.09\frac{a}{b}\frac{\lambda g}{\lambda}\cos^2\left(\frac{\pi}{2}\frac{\lambda}{\lambda g}\right)\sin^2\left(\frac{\pi\delta}{a}\right), \qquad (1)$$

where a and b are the internal guide dimensions, λ is the free-space wavelength, λg the wavelength of the dominant mode in the guide, and δ the offset of the center line of the slot from the center line of the broad face of the guide.

In the derivation of the above result it is assumed that the guide walls are perfectly conducting and of negligible thickness, that the guide transmits the TE_{10}-mode only, and that the width of the slot is small compared with its length. Furthermore, the power radiated by the slot for a given excitation across the slot is calculated as if the slot were situated in an infinite conducting plane. Although the last assumption may be seriously at fault, experimental results are in very good agreement with Eq. (1).

If the slot radiates into a parallel-plate region attached to the wave guide instead of into free space, the last assumption is no longer necessary. The field between the parallel plates can be expanded in normal modes and a rigorous expression can be found for the power radiated from the slot. The resultant formula for the conductance of the slot should therefore be quite accurate.

[1] H. Gruenberg, "Second-order beams of slotted waveguide arrays" (to be published).
[2] D. W. Fry and F. K. Goward, *Aerials for Centimetre Wave-Lengths* (Cambridge University Press, Cambridge, 1950), pp. 131–133.
[3] A. F. Stevenson, J. Appl. Phys. **19**, 24–28 (1948).

ANALYSIS

It is clear from Stevenson's[3] paper that the derivation for the conductance of a resonant slot will follow similar lines, whether the slot is radiating into free space, into a parallel-plate region, or any other confined space. The only difference will be in the contribution to the slot impedance from the fields external to the wave guide. The slot conductance at resonance is inversely proportional to the power radiated from the slot. If P_0 is the power radiated from a slot into the half-space bounded by an infinite conducting plane, and P the power radiated into an unbounded parallel-plate region for the same excitation voltage across the slot, then the corresponding slot conductances are related by

$$g/g_0=P_0/P. \qquad (2)$$

The power radiated by a half-wave slot in an infinite conducting plane can be calculated in the usual manner by integrating the Poynting flux over an infinite hemisphere on one side of the plane, or by making use of Babinet's principle and the corresponding result for a wire antenna.[4] The result is, using rationalized M.K.S. units,

$$P_0=\frac{V_0^2}{120\pi^2}\int_0^{\pi/2}\frac{\cos^2(\frac{1}{2}\pi\cos\theta)}{\sin\theta}d\theta$$

$$=\frac{V_0^2}{480\pi^2}[\ln 2\pi\gamma-Ci(2\pi)]=\frac{V_0^2}{1940}. \qquad (3)$$

It is assumed that the slot is excited by a transverse electric field with a maximum voltage V_0 at the center

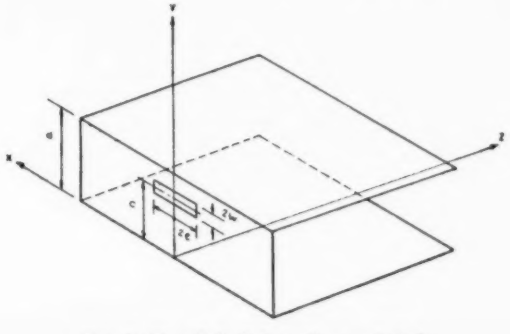


FIG. 1. Parallel-plate region and slot.

[4] N. A. Begovich, Proc. Inst. Radio Engrs. **38**, 803–806 (1950).

of the slot tapering sinusoidally to zero at the ends of the slot.

The power radiated from the slot into a parallel-plate region can be found by expanding the field between the plates in normal modes and integrating the Poynting flux over the area of the slot. The geometry of the system is shown in Fig. 1. The parallel-plate region consists of two conducting plates at a distance d apart. The region is bounded at $z=0$ by the face of the wave guide containing the slot, but is assumed to extend to infinity in all other directions. In practice it may be terminated by an absorbing load such as a radiating horn aperture.

The voltage distribution along the slot will again be assumed sinusoidal, so that the tangential electric field in the plane $z=0$ may be described by

$$\left.\begin{aligned} &E_y(0)=(V_0/2w)\cos(\pi x/2l) \\ &\qquad \text{for} \quad -l<x<+l, \quad c-w<y<c+w \\ &E_y(0)=0 \text{ elsewhere.} \end{aligned}\right\} \tag{4}$$

This field can be written in terms of a double Fourier expansion

$$E_y(0)=\sum_{n=0}^{\infty} A_n \cos\frac{n\pi y}{d}\int_0^{\infty} B(k)\cos kx\,dk, \tag{5}$$

where

$$\left.\begin{aligned} &A_0=V_0/d \\ &A_n=\frac{2V_0}{d}\cos\frac{n\pi c}{d}\frac{\sin(n\pi w/d)}{(n\pi w/d)} \quad \text{for} \quad n\neq 0 \\ &B(k)=\frac{l\cos kl}{(\pi/2)^2-(kl)^2}. \end{aligned}\right\} \tag{6}$$

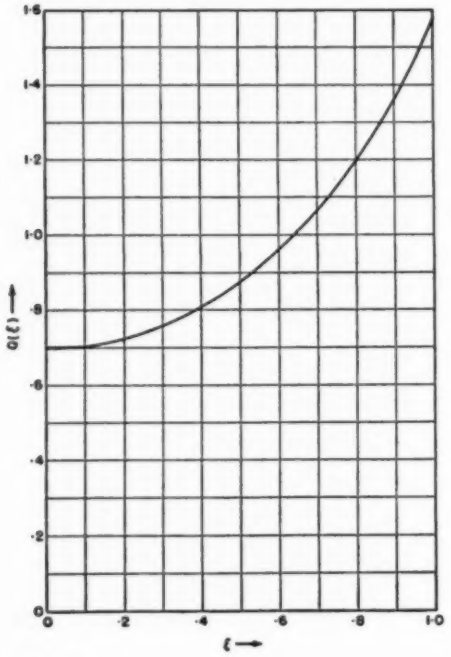


FIG. 2. Plot of
$Q(\xi)=\int_0^{\pi/2}\{\cos^2[(\pi/2)(1-\xi^2)^{\frac{1}{2}}\sin\theta]/[1-(1-\xi^2)\sin^2\theta]\}d\theta.$

These relations together with the vanishing of the tangential electric field on the parallel plates and the restriction to outgoing waves at infinity specify the complete set of boundary conditions for the problem.

By inspection, a solution of the wave equation within the parallel-plate region satisfying all boundary conditions is

$$\left.\begin{aligned} &E_y(z)=\sum_{n=0}^{\infty} A_n\cos\frac{n\pi y}{d}\int_0^{\infty} B(k)\cos kx\,e^{-\Gamma_n(k)z}dk \\ &E_z(z)=-\sum_{n=0}^{\infty}\frac{n\pi}{d}A_n\sin\frac{n\pi y}{d}\int_0^{\infty}\frac{B(k)}{\Gamma_n(k)} \\ &\qquad\qquad \times\cos kx\,e^{-\Gamma_n(k)z}dk \\ &E_x(z)=0, \end{aligned}\right\} \tag{7}$$

if

$$\Gamma_n^2(k)=k^2+(n\pi/d)^2-\beta^2, \quad \beta=2\pi/\lambda. \tag{8}$$

It is easily verified that this field also satisfies Maxwell's divergence relation and therefore represents *the* unique field solution.

To obtain the power flow through the slot one must know the component of the magnetic field in the plane of the slot, perpendicular to and in phase with the electric field given by (4) or (5). By Maxwell's curl relation and Eqs. (7) and (8) one gets, at $z=0$,

$$\begin{aligned} H_x(0)&=\frac{j}{\mu\omega}\left(\frac{\partial E_z}{\partial y}-\frac{\partial E_y}{\partial z}\right)_{z=0} \\ &=\frac{j}{\omega\mu}\sum_{n=0}^{\infty} A_n\cos\frac{n\pi y}{d}\int_0^{\infty}\frac{k^2-\beta^2}{\Gamma_n(k)}B(k)\cos kx\,dk. \end{aligned} \tag{9}$$

The contributions to $H_x(0)$ from all the nonpropagating modes ($\Gamma_n>0$) are seen to be out of phase with $E_y(0)$. This must be so, of course, since the nonpropagating modes cannot contribute anything to the average power flow. From (8), propagation can only occur when

$$0<k<[\beta^2-(n\pi/d)^2]^{\frac{1}{2}} \tag{10}$$

and

$$\beta=2\pi/\lambda>n\pi/d. \tag{11}$$

The case in which only the principal mode ($n=0$) is propagated will be considered first. To achieve this, one must keep $d<\lambda/2$ and $0<k<\beta$. The in-phase component of $H_x(0)$ is then

$$H_x^{\,i}(0)=-(A_0/\omega\mu)\int_0^{\beta}(\beta^2-k^2)^{\frac{1}{2}}B(k)\cos kx\,dk. \tag{12}$$

The average power flow through the slot is

$$P=-\frac{1}{2}\int_{c-w}^{c+w}\int_{-l}^{+l} E_y(0)H_x^{\,i}(0)\,dx\,dy. \tag{13}$$

Substituting from (4), (12), and (6) and noting that $\beta l=\pi/2$ for a half-wave slot, one obtains after integration with respect to x and y:

$$P=\frac{2V_0^2}{\pi\omega\mu d}\int_0^{\pi/2}\frac{\cos^2(\frac{1}{2}\pi\sin\theta)}{\cos^2\theta}d\theta, \tag{14}$$

where kl has been replaced by a new variable $(\pi/2)\sin\theta$. The above integral can be evaluated numerically and is equal to 0.701. The factor $\mu\omega$ can be written as the product of the intrinsic impedance of the medium, $(\mu/\epsilon)^{\frac{1}{2}}=120\pi$ ohms, and the phase constant $\beta=\omega(\mu\epsilon)^{\frac{1}{2}}=2\pi/\lambda$. Hence

$$P=\frac{0.701V_0^2}{120\pi^2}\frac{\lambda}{d}=\frac{V_0^2}{5330}\frac{\lambda}{d}, \tag{15}$$

and from (2), (3), and (15)

$$g/g_0=2.75d/\lambda \quad \text{for} \quad d<\lambda/2. \tag{16}$$

The conductance is seen to be independent of the location of the slot with respect to the parallel plates. It varies linearly with the plate spacing provided that $d<\lambda/2$ and only the principal mode is propagated. There is a plate spacing $(d=\lambda/2.75)$ for which the slot conductance is the same as that for the slot radiating from the wave guide into free space. Clearly the same condition must be approached again as $d\to\infty$ and all the higher order modes contribute to the slot conductance. This will be shown in the next section.

Equation (16) can be extended to hold for any plate spacing. Let d be such that $n=N$ is the highest mode that can be propagated, i.e.,

$$N\lambda/2<d<(N+1)\lambda/2. \tag{17}$$

The continuous mode number k is restricted as in (10) before. The in-phase component of the magnetic field (12) must be replaced by

$$\left.\begin{aligned} H_z^i(0)&=-\sum_{n=0}^{N}\frac{A_n}{\omega\mu}\cos\frac{n\pi y}{d}\\ &\quad\times\int_0^{\beta_n}\frac{\beta^2-k^2}{(\beta_n^2-k^2)^{\frac{1}{2}}}B(k)\cos kx\,dk\\ \beta_n&=[\beta^2-(n\pi/d)^2]^{\frac{1}{2}}. \end{aligned}\right\} \tag{18}$$

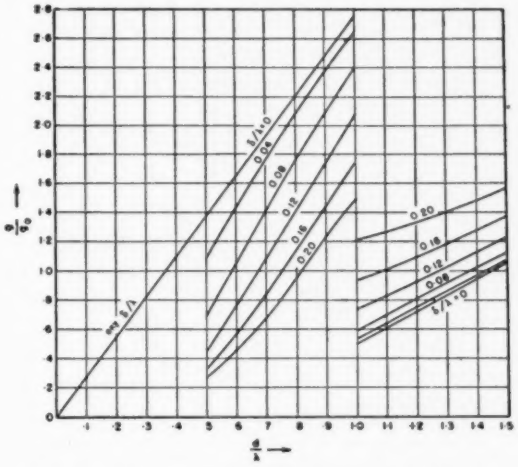


FIG. 3. Plot of conductance ratio as a function of plate spacing and slot offset.

Proceeding as before one obtains for the power radiated

$$P=\frac{V_0^2}{120\pi^2}\frac{\lambda}{d}Q(0)\left\{1+2\sum_{n=1}^{N}\cos^2\frac{n\pi c}{d}\times\left[\frac{\sin(n\pi w/d)}{n\pi w/d}\right]^2\frac{Q(n\lambda/2d)}{Q(0)}\right\}, \tag{19}$$

where

$$Q(\xi)=\int_0^{\pi/2}\frac{\cos^2[(\pi/2)(1-\xi^2)^{\frac{1}{2}}\sin\theta]}{1-(1-\xi^2)\sin^2\theta}d\theta. \tag{20}$$

The value of the integral $Q(\xi)$ is plotted in Fig. 2. Note that $Q(0)=0.701$ is identical with the integral in (14). Because of the inequality (16), the factor $\sin(n\pi w/d)/(n\pi w/d)$ is nearly unity in all practical cases where the width of the slot is small compared with the wavelength. Equation (16) is then modified to

$$g/g_0=\frac{2.75d/\lambda}{1+2\sum_{n=1}^{N}\cos^2(n\pi c/d)Q(n\lambda/2d)/Q(0)} \quad \text{for} \quad d>\lambda/2. \tag{21}$$

If the parallel-plate region is centered with respect to the center line of the face of the wave guide containing the slot, then $c=d/2+\delta$, and

$$g/g_0=\frac{2.75d/\lambda}{1+\sum_{n=1}^{N}\{1+(-1)^n\cos[4\pi(n\lambda/2d)(\delta/\lambda)]\}Q(n\lambda/2d)/Q(0)} \quad \text{for} \quad d>\lambda/2. \tag{22}$$

DISCUSSION

It is worthwhile to examine the expressions for g/g_0 given in (16) and (22) more closely. The conductance ratio is plotted in Fig. 3 as a function of d/λ for various values of δ/λ. The curve for zero offset $(\delta=0)$ is, of course, trivial, since the slot does not radiate in this case and both g_0 and g are zero. For plate spacings up to half a wavelength, g/g_0 varies linearly with d and is

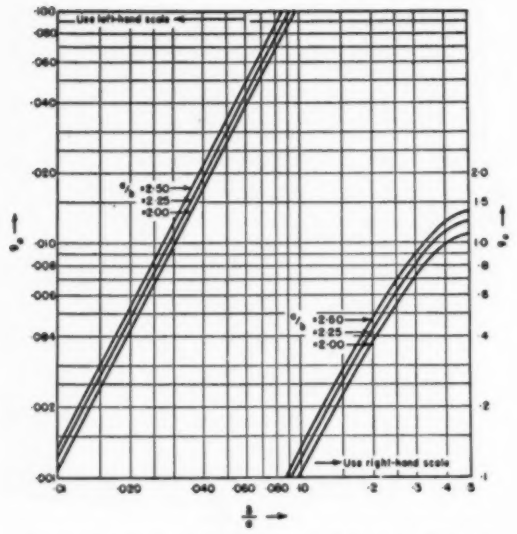


FIG. 4. Conductance of resonant slots radiating into free space at $\lambda_0=0.7\lambda_c=1.4a$.

independent of the slot offset as was pointed out before. The physical reason for this is that only the principal mode is propagated, the amplitude of which is independent of the slot offset. As d is increased, higher modes can be propagated. Their amplitudes depend on the slot offset so that we may expect different curves for different values of δ.

As d increases to infinity we must expect g/g_0 to approach unity independent of the value of δ. To prove this, it is only necessary to show that for $2w \ll \lambda$ and $d \to \infty$, the expression (19) for P approaches P_0 as given by (3). The first term in (19) approaches zero as $d \to \infty$, while the summation becomes

$$\lim_{d\to\infty} P = \frac{V_0^2}{60\pi^3} \lim_{N\to\infty} \sum_{\text{even } n}^{\infty} (2/N)Q(n/N), \tag{23}$$

because $c/d \to \frac{1}{2}$ and $2d/\lambda \to N$ by (17), and because it was assumed that $2w \ll \lambda$.

By the fundamental theorem of integral calculus

$$\lim_{N\to\infty} \sum_{\text{even } n}^{\infty} (2/N)Q(n/N) = \int_0^1 Q(\xi)d\xi$$

$$= \int_{\xi=0}^{\xi=1} \int_{\theta=0}^{\theta=\pi/2} \frac{\cos^2[(\pi/2)(1-\xi^2)^{\frac{1}{2}} \sin\theta]d\theta d\xi}{1-(1-\xi^2)\sin^2\theta}. \tag{24}$$

This double integral can be evaluated by substituting u for $(1-\xi^2)^{\frac{1}{2}}\sin\theta$ and v for $(1-\xi^2)^{\frac{1}{2}}$. The Jacobian of the transformation is

$$\partial(\xi,\theta)/\partial(u,v) = [(1-u^2/v^2)(1-v^2)]^{-\frac{1}{2}}, \tag{25}$$

and the rectangular domain of integration bounded by the straight lines $\xi=0$, $\xi=1$, $\theta=0$, and $\theta=\pi/2$ is transformed into the triangular domain bounded by $u=0$, $v=u$, and $v=1$. The double integral becomes

$$\int_{u=0}^{u=1} \int_{v=u}^{v=1} \frac{v\cos^2(\pi u/2)dudv}{(1-u^2)[(v^2-u^2)(1-v^2)]^{\frac{1}{2}}}. \tag{26}$$

The integration with respect to v is carried out by introducing the new variable $t=[(1-v^2)/(1-u^2)]^{\frac{1}{2}}$ and leads to a value of $\pi/2$. Hence

$$\lim_{d\to\infty} P = \frac{V_0^2}{120\pi^2} \int_0^1 \frac{\cos^2(\pi u/2)du}{1-u^2}. \tag{27}$$

This is seen to be identical with (3) by substituting $\cos\theta$ for u.

Striking features of Fig. 3 are the discontinuities in the conductance ratio at plate spacings of multiples of half-wavelengths and the inversion of the curves at $d=\lambda$. This may appear strange at first sight. It must be remembered, however, that (22) was derived on the assumption of perfectly conducting boundaries. Near cutoff of the various modes this is not quite justified, since the finite conductivity modifies the propagation constant there. In practice the conductance curve will thus be continuous with the sharp corners rounded off. Nevertheless, the conductance must change very rapidly for very small changes in plate spacing near these resonances. This has been verified by measurements in standard RG50/U guide (1.500-inch by 0.750-inch outside dimensions, 0.064-inch wall). These abrupt changes in the slot conductance also explain the fact that measurements made with a plate spacing near half a wavelength are often very erratic and appear to be inconsistent.

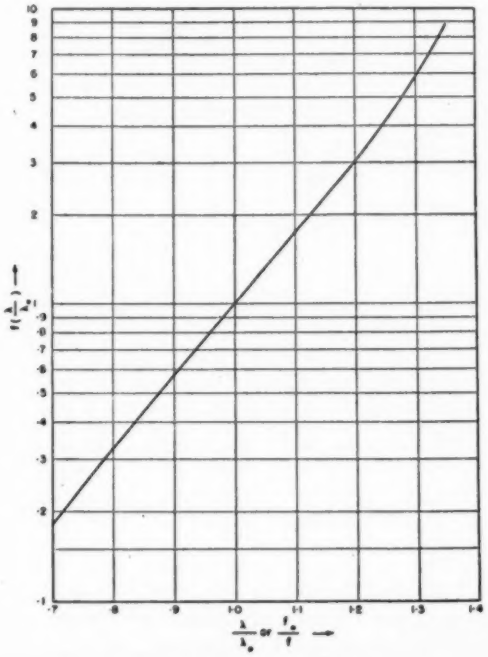


FIG. 5. Plot of frequency correction factor.

Measurements were made at plate spacings corresponding to values of d/λ between 0.4 and 0.9. The parallel-plate region was matched to free space by a wooden wedge coated with Aquadac. The slots were rectangular and $\frac{3}{16}$ inch wide. Although the resonant frequency of the slots varied somewhat with plate spacing, the measured conductances were in excellent agreement with the above theory. The measured values of g_0 were approximately 4 percent higher than those calculated from Stevenson's formula. This discrepancy is presumably due to the assumptions used in arriving at Eq. (1). It is believed that the values of g obtained from (1), (16), and (22) agree more closely with experiment because of the reasons outlined in the introduction.

For completeness, universal curves are included for the free-space conductance g_0 calculated from Stevenson's formula (1). Figure 4 gives the value of g_0 for a/b ratios common in standard wave guide at a wavelength λ_0 corresponding to 70 percent of the cut-off wavelength. Figure 5 gives a correction multiplier to obtain g_0 at other wavelengths.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

On the Evaluation of Noise Samples

ARNOLD J. F. SIEGERT*
The RAND Corporation, Santa Monica, California
(Received March 3, 1952)

Unless the primary source of noise is one of those which is theoretically tractable, the statistical properties of the noise have to be inferred from samples. We have developed some criteria to aid in the decision whether a sample can reasonably be assumed to have come from a Gaussian noise with predetermined parameters.

INTRODUCTION

THE optimum design of instruments such as aiming devices requires the knowledge of the statistical properties of the noise which will accompany the signals to which the devices react. Unless the primary source of this noise is one of those which is theoretically completely tractable, such as thermal noise, shot effect, etc., the statistical properties of the noise must be obtained empirically from samples. Since the size of the sample is often limited by the cost of obtaining samples, the problem arises of the extent to which the statistical properties obtained from a given sample can be relied upon. This problem has been given precise formulation in the statistical literature for samples taken from populations consisting of discrete elements. Since the elements of the sample in our case are the observed values of $y(t)$ in a time interval $(0, T)$, they are neither discrete nor statistically independent. The sample estimates are functionals of $y(t)$ and the only nontrivial functionals for which—even for a Gaussian random function—the probability distribution is at present known or can be obtained are the linear integral forms[1] (weighted sample means) and the quadratic integral forms[2] (such as mean square, correlation function, and spectrum of the sample).[3] The expectation value and average square of sample estimates can usually be obtained for Gaussian random functions and yield some information about the probability distribution through the Bienaymé-Tchebycheff inequality.[4]

Since most of the above results have been calculated for Gaussian random functions only, it seemed of interest to develop the basis for a criterion by which one may judge whether at least the first distribution of the sample is sufficiently close to a Gaussian distribution so that the sample could have come from a Gaussian random function. We define the first distribution $\vartheta(a, T)$ of a sample $y(t)$ as that fraction of the interval $(0, T)$ during which $y(t) > a$. We have computed the fluctuation $S^2 = \langle [\vartheta(a, T) - \phi(a)]^2 \rangle_{Av}$ and the integrals $\mathcal{S}$ and $\mathcal{S}_2$ of the fluctuation over all values of a without weighting and with a weight function $G(a)$, respectively, for Gaussian random functions with first probability distribution $\phi(a)$ and with arbitrarily given correlation function.

The hypothesis that the sample came from a Gaussian distribution with specified parameters can be rejected if the deviation $|\vartheta(a, T) - \phi(a)|$ at a predetermined[5] value a appreciably exceeds the value S, or if the integral

$$\int_{-\infty}^{\infty} [\vartheta(a, T) - \phi(a)]^2 (G(a)) da$$

---

* Consultant to Rand Corporation; permanently at the Department of Physics, Northwestern University, Evanston, Illinois.

[1] N. Wiener, Acta Math. **55**, 117 (1930).

[2] M. Kac and A. J. F. Siegert, J. Appl. Phys. **18**, 383 (1947) and Ann. Math. Stat. **18**, 438 (1947).

[3] For Markoffian random functions the probability distribution of the largest value in the sample can be obtained [A. J. F. Siegert, Phys. Rev. **81**, 617 (1951)] and for an especially simple Markoffian random function the Laplace transform of the distribution of the integral of the absolute value and certain other nonlinear functionals are known [M. Kac, Trans. Am. Math. Soc. **59**, 401 (1946); **65**, 1 (1949)].

[4] See Harald Cramér, *Mathematical Methods of Statistics* (Princeton University Press, Princeton, New Jersey 1946), pp. 182–183.

[5] "Predetermined" means that the value of a is chosen without knowledge of the sample, and not, for instance, at the point for which $|\vartheta(a, T) - \phi(a)|$ has its largest value.

appreciably exceeds S or S_2, respectively. More precisely we know from the Bienaymé-Tchebycheff inequality

$$\text{prob}\{|\vartheta(a,T)-\phi(a)|\geq kS\}\leq 1/k^2$$

for any predetermined value of a, and

$$\text{prob}\left\{\int_{-\infty}^{\infty}(G(a))[\vartheta(a,T)-\phi(a)]^2da\geq kS_{(2)}\right\}\leq 1/k.$$

There remains, of course, as always, an arbitrariness in choosing a value for the probability at which one wishes to reject the hypothetical distribution, rather than accept an improbable outcome of the experiment.

(1) Definitions

The fraction of the time T during which $y(t)>a$ is denoted by $\vartheta(a,T)$ and expressed as

$$\vartheta(a,T)=\frac{1}{T}\int_0^T\beta(y(t)-a)dt, \tag{1.1}$$

where β is defined by

$$\beta(x)=\begin{matrix}0\\1\end{matrix}\quad\text{for}\quad\begin{matrix}x\leq 0\\x>0\end{matrix}. \tag{1.2}$$

Not only the problem of finding the probability distribution of $\vartheta(a,T)$, but even the special cases $\text{prob}\{\vartheta(a,T)=0\}$ and $\text{prob}\{\vartheta(a,T)=1\}$ seem to be extremely difficult, except for Markoffian random functions, since their solution is closely related to the solution of the first passage time probability problem.

We define the first and second probability functions of the stationary random function $y(t)$ by

$$\text{prob}\{y_1\leq y(t_1)<y_1+dy_1\}=W(y_1)dy_1 \tag{1.3}$$

and

$$\text{prob}\begin{Bmatrix}y_1\leq y(t_1)<y_1+dy_1\\y_2\leq y(t_2)<y_2+dy_2\end{Bmatrix}=W_2(y_1,y_2;t_2-t_1)dy_1dy_2, \tag{1.4}$$

and obtain

$$\langle\vartheta(a,T)\rangle_{Av}=\frac{1}{T}\int_0^T\langle\beta(y(t)-a)\rangle_{Av}dt=\phi(a),$$

where

$$\phi(a)\equiv\int_a^{\infty}W(y)dy \tag{1.5}$$

is the probability that $y(t)$ exceeds the value a. For the second moment we get

$$\langle\vartheta^2(a,T)\rangle_{Av}\equiv\left\langle\left(\frac{1}{T}\int_0^T\beta(y(t)-a)dt\right)^2\right\rangle_{Av}$$

$$=\frac{1}{T^2}\int_0^T\int_0^T dt_1dt_2\langle\beta(y(t_1)-a)\beta(y(t_2)-a)\rangle_{Av}$$

$$=\frac{1}{T^2}\int_0^T\int_0^T dt_1dt_2\int_a^{\infty}\int_a^{\infty}W_2(y_1,y_2;t_2-t_1)dy_1dy_2$$

$$=\frac{2}{T}\int_0^T dt\left(1-\frac{t}{T}\right)\int_a^{\infty}\int_a^{\infty}W_2(y_1,y_2;t)dy_1dy_2. \tag{1.6}$$

(2) Rough Approximations

For orientation we consider first the oversimplified[6] case

$$W_2(y_1,y_2;t_2-t_1)=\begin{cases}W(y_1)\delta(y_2-y_1) & \text{for}\quad|t_2-t_1|\leq\tau\\W(y_1)W(y_2) & \text{for}\quad|t_2-t_1|>\tau\end{cases}. \tag{2.1}$$

A simple geometric consideration yields for this case

$$\langle\vartheta^2(a,T)\rangle_{Av}=\left(1-\frac{\tau}{T}\right)^2\phi^2(a)+\left(1-\left(1-\frac{\tau}{T}\right)^2\right)\phi(a), \tag{2.2}$$

so that the fluctuation becomes

$$\langle\vartheta^2(a,T)\rangle_{Av}-\langle\vartheta(a,T)^2{}_{Av}$$

$$=\left(1-\left(1-\frac{\tau}{T}\right)^2\right)[\phi(a)-\phi^2(a)]$$

$$=2\frac{\tau}{T}\left(1-\frac{\tau}{2T}\right)[\phi(a)-\phi^2(a)]. \tag{2.3}$$

As was to be expected the fluctuation vanishes for $\tau/T\to 0$ and becomes equal to $\phi(a)-\phi^2(a)$ for $\tau/T\to 1$.

A rough idea of the distribution of $\vartheta(a,T)$ can be obtained by dividing the interval T into m equal parts, choosing m such that the values of $y(t)$ at the center-points t_j of the intervals can be considered as approximately independent random variables. A set of inde-

[6] This is not a realistic assumption, since it yields for the normalized correlation function

$$\rho(t_2-t_1)\equiv(\langle y(t_2)y(t_1)\rangle_{Av}-\langle y\rangle_{Av}{}^2)/(\langle y^2\rangle_{Av}-\langle y\rangle_{Av}{}^2)$$

the values

$$\rho(t)=\begin{cases}1 & \text{for}\quad|t|\leq\tau\\0 & \text{for}\quad|t|>\tau\end{cases}.$$

This is not a possible correlation function, since its Fourier transform can assume negative values.

pendent random variables ϑ_j is then defined such that $\vartheta_j=0$ or 1 if $\beta(y(t_j)-a)=0$ or 1, respectively.

The mean

$$\vartheta_{(m)}=\frac{1}{m}\sum_{j=1}^{m}\vartheta_j$$

is an approximation for $\vartheta(a, T)$.

The probability distribution for $\vartheta_{(m)}$ is then

$$\text{prob}\left\{\vartheta_{(m)}=\frac{n}{m}\right\}=\binom{m}{n}\phi^n(a)[1-\phi(a)]^{m-n} \quad (2.4)$$

with mean $\phi(a)$ and fluctuation $S^2=m^{-1}(\phi-\phi^2)$. This is in agreement with Eq. (2.3) for $\tau/T\ll 1$.

(3) Calculation of $<\vartheta^2(a,T)>_{\text{Av}}$ for Gaussian Random Functions

For the Gaussian random function we have

$$W(y)=\frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left(-\frac{y^2}{2\sigma^2}\right), \quad (3.1)$$

$$W_2(y_1, y_2, t)=\frac{1}{2\pi\sigma^2(1-\rho^2)^{\frac{1}{2}}}\times\exp\left(-\frac{y_1^2+y_2^2-2\rho y_1y_2}{2\sigma^2(1-\rho^2)}\right). \quad (3.2)$$

To evaluate the integral in Eq. (1.6) we expand W_2 using the equation[7]

$$\sum_{n=0}^{\infty}\frac{H_n(x)H_n(y)}{n!}\rho^n=\frac{1}{(1-\rho^2)^{\frac{1}{2}}}\times\exp\left(-\frac{\rho^2(x^2+y^2)-2\rho xy}{2(1-\rho^2)}\right), \quad (3.3)$$

and obtain

$$W_2(y_1, y_2, t)=\frac{1}{2\pi\sigma^2(1-\rho^2)^{\frac{1}{2}}}\exp\left\{-\left[\frac{y_1^2+y_2^2}{2\sigma^2}+\frac{\rho^2(y_1^2+y_2^2)-2\rho y_1y_2}{2\sigma^2(1-\rho^2)}\right]\right\}$$

$$=\frac{1}{2\pi\sigma^2}\exp\left(-\frac{y_1^2+y_2^2}{2\sigma^2}\right)\times\sum_{n=0}^{\infty}\frac{H_n(y_1/\sigma)H_n(y_2/\sigma)}{n!}\rho^n, \quad (3.4)$$

where $H_n(x)$ are the Hermite polynominals. With

$$\int_a^{\infty}\frac{dy}{\sigma}\exp\left(-\frac{y^2}{2\sigma^2}\right)H_n\left(\frac{y}{\sigma}\right)$$

$$=\exp\left(-\frac{a^2}{2\sigma^2}\right)H_{n-1}(a/\sigma) \quad \text{for} \quad n>0$$

$$=(2\pi)^{\frac{1}{2}}\phi(a) \quad \text{for} \quad n=0, \quad (3.5)$$

we get

$$\int_a^{\infty}\int_a^{\infty}W_2(y_1, y_2; t)dy_1dy_2$$

$$=\phi^2(a)+\frac{1}{2\pi}\exp\left(-\frac{a^2}{\sigma^2}\right)\sum_{n=1}^{\infty}\frac{H_{n-1}^2(a/\sigma)}{n!}\rho^n$$

$$=\phi^2(a)+\frac{1}{2\pi}\exp\left(-\frac{a^2}{\sigma^2}\right)\int_0^{\rho}\sum_{n=1}^{\infty}\frac{H_{n-1}^2(a/\sigma)}{(n-1)!}r^{n-1}dr$$

$$=\phi^2(a)+\exp\left(-\frac{a^2}{\sigma^2}\right)\frac{1}{2\pi}\int_0^{\rho}\sum_{n=0}^{\infty}\frac{H_n^2(a/\sigma)}{n!}r^ndr$$

$$=\phi^2(a)+\frac{1}{2\pi}\int_0^{\rho}\frac{dr}{(1-r^2)^{\frac{1}{2}}}\exp\left(-\frac{a^2}{\sigma^2}\frac{1}{1+r}\right), \quad (3.6)$$

and

$$S^2(a, t)\equiv\langle\vartheta^2(a, T)\rangle_{\text{Av}}-\langle\vartheta(a, T)\rangle^2_{\text{Av}}$$

$$=\frac{1}{\pi T}\int_0^T\left(1-\frac{t}{T}\right)dt\int_0^{\rho(t)}\frac{dr}{(1-r^2)^{\frac{1}{2}}}\times\exp\left(-\frac{a^2}{\sigma^2}\frac{1}{1+r}\right). \quad (3.7)$$

S^2 can be written in terms of single integrals in the form

$$S^2=\frac{1}{2\pi}\int_0^{\rho(T)}\frac{dr}{(1-r^2)^{\frac{1}{2}}}\exp\left[-\frac{a^2}{\sigma^2(1+r)}\right]$$

$$-\frac{1}{\pi T}\int_0^T t\left(1-\frac{t}{2T}\right)\frac{d\rho}{dt}\frac{1}{(1-\rho^2)^{\frac{1}{2}}}\times\exp\left(-\frac{a^2}{\sigma^2}\frac{1}{1+\rho}\right)dt. \quad (3.8)$$

Usually, $\rho(T)$ will be so small that the first integral can be neglected. With the oversimplified form of $\rho(t)$

$$\rho(t)=\begin{cases}1 & \text{for} \quad |t|\leq\tau\\ 0 & \text{for} \quad |t|>\tau,\end{cases} \quad (3.9)$$

[7] H. Cramér, reference 4, p. 133.

we get

$$S_1^2 \equiv \frac{1}{\pi T}\int_0^\tau \left(1-\frac{t}{T}\right)dt \int_0^1 \frac{dr}{(1-r^2)^{\frac{1}{2}}} \times \exp\left(-\frac{a^2}{\sigma^2}\frac{1}{1+r}\right)$$

$$= \frac{2\tau}{T}\left(1-\frac{\tau}{2T}\right)\cdot\frac{1}{2\pi}\int_0^1 \frac{dr}{(1-r^2)^{\frac{1}{2}}} \times \exp\left(-\frac{a^2}{\sigma^2}\frac{1}{1+r}\right). \quad (3.10)$$

The identity

$$\frac{1}{2\pi}\int_0^1 \frac{dr}{(1-r^2)^{\frac{1}{2}}}\exp\left(-\frac{a^2}{\sigma^2(1+r)}\right) = \phi(a)-\phi^2(a)$$

is easily verified.

A convenient expansion for $a \ll \sigma$ is obtained by changing to the variable

$$y = [(1-r)/(1+r)]^{\frac{1}{2}}, \quad (3.11)$$

and writing

$$\int_0^{r} \frac{dr}{(1-r^2)^{\frac{1}{2}}}\exp\left(-\frac{a^2}{\sigma^2(1+r)}\right) = 2\int_{(1-\rho/1+\rho)^{\frac{1}{2}}}^1 \frac{dy}{1+y^2}\exp\left(-\frac{(1+y^2)a^2}{2\sigma^2}\right). \quad (3.12)$$

We get

$$S^2(a,T) \cong \frac{1}{\pi T}\exp\left(-\frac{a^2}{2\sigma^2}\right) \times \int_0^T \left[\left(1+\frac{a^2}{2\sigma^2}\right)\arcsin\rho - \frac{a^2}{\sigma^2}\left(1-\left(\frac{1-\rho}{1+\rho}\right)^{\frac{1}{2}}\right)\right]\left(1-\frac{t}{T}\right)dt, \quad (3.13)$$

where terms of order $(a/\sigma)^4$ have been neglected under the integral.

To judge the sample distribution in its entirety it may also be useful to compute from the data the quantity

$$S^* \equiv \int_{-\infty}^{\infty} [\vartheta(a,T)-\phi(a)]^2 da \quad (3.14)$$

and compare it with its expected value S. We have

$$S = \langle S^* \rangle_{Av} = \frac{1}{\pi T}\int_0^T \left(1-\frac{t}{T}\right)dt \times \int_0^{\rho(t)} \frac{dr}{(1-r^2)^{\frac{1}{2}}}[\pi\sigma^2(1+r)]^{\frac{1}{2}}$$

$$= \frac{\sigma}{T(\pi)^{\frac{1}{2}}}\int_0^T \left(1-\frac{t}{T}\right)dt \int_0^{\rho(t)} \frac{dr}{(1-r)^{\frac{1}{2}}}$$

$$= \frac{2\sigma}{T(\pi)^{\frac{1}{2}}}\int_0^T \left(1-\frac{t}{T}\right)dt\,(1-[1-\rho(t)]^{\frac{1}{2}}). \quad (3.15)$$

For $\rho(t) = e^{-\beta|t|}$ (Markoffian Gaussian random function) the integral has been evaluated in Appendix I, neglecting terms smaller than $[2\sigma/\beta T(\pi)^{\frac{1}{2}}]e^{-\beta T}$. The result is

$$S \cong \frac{2\sigma}{\beta T(\pi)^{\frac{1}{2}}}(0.61-0.54(\beta T)^{-1}) \quad (3.16)$$

in this approximation.

Another test which places more emphasis on the deviations of the sample distribution from the Gaussian at large values of $|a/\sigma|$ is obtained by computing from the data the quantity[8]

$$S_2^* = \int_{-\infty}^{\infty} [\vartheta(a,T)-\phi(a)]^2 \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da. \quad (3.17)$$

Denoting by S_2 the expected value of this quantity we have

$$S_2 = \frac{1}{\pi T}\int_0^T \left(1-\frac{t}{T}\right)dt \int_0^{r} \frac{dr}{(1-r^2)^{\frac{1}{2}}} \times \int_{-\infty}^{\infty} \frac{da}{(2\pi\sigma^2)^{\frac{1}{2}}}\exp\left[-\frac{a^2}{2\sigma^2}\left(\frac{2}{1+r}-r\right)\right]$$

$$= \frac{1}{\pi T}\int_0^T \left(1-\frac{t}{T}\right)dt \int_0^{r} \frac{dr}{(1-r^2)^{\frac{1}{2}}}\left(\frac{1+r}{1-r}\right)^{\frac{1}{2}}$$

$$= -\frac{1}{\pi T}\int_0^T \left(1-\frac{t}{T}\right)dt\, lg[1-\rho(t)]. \quad (3.18)$$

For the Markoffian case we have (neglecting terms of order $e^{-\beta T}$) the result

$$S_2 = \pi^{-1}(\beta T)^{-1}[\zeta(2)-(\beta T)^{-1}\zeta(3)], \quad (3.19)$$

[8] Approximative formulas for the contributions to this integral from the regions in which $\vartheta(a,T)=0$ or 1 are given in Appendix II.

(see Appendix III) with $\zeta(2)=(\pi^2/6)=1.65$ and $\zeta(3)=1.20$.

(4) Discussion

The results of our calculations state in various forms how far the empirical first distributions of samples taken from a Gaussian noise with specified parameters will deviate from the true first distribution of this Gaussian noise. They can thus be used to judge whether a sample of unknown source can reasonably be assumed to have come from a Gaussian noise with predetermined parameters. Our results cannot in all cases be used to judge whether the sample can reasonably be assumed to have come from a Gaussian noise with parameters obtained from the sample itself. To obtain criteria for the latter decision would require the calculation of quantities such as

$$\langle[\vartheta(a,T)-\phi(a,\mu^*,\sigma^*)]^2\rangle_M, \tag{4.1}$$

where $\phi(a,\mu^*,\sigma^*)$ is the Gaussian distribution with mean μ^* and standard deviation σ^*, which are both computed from—and thus functionals of—the sample function $y(t)$. Expectation values such as (4.1) are much more difficult to calculate than our averages and may not be obtainable in a useful form.

An answer to the second question above can, however, still be obtained from our results, if the size T of the sample is sufficiently large, such that its mean μ^* and standard deviation σ^* become reliable estimates of the true mean μ and standard deviation σ. In the computation of $[\vartheta(a,T)-\phi(a)]^2$, $\mathcal{S}^*$ and $\mathcal{S}_2^*$ from the sample function, one can then use various combinations of values μ and σ deviating from μ^* and σ^* rsp by their respective expected errors (which are not difficult to estimate). If the various values of $[\vartheta(a,T)-\phi(a)]^2$, $\mathcal{S}^*$ and $\mathcal{S}_2^*$ thus computed all lead to the same conclusion, our results apply to the second question. We believe that the question whether a sample can have come from a Gaussian noise can reasonably be asked only in this case, when at least the estimates of the mean and standard deviation obtained from the sample are sufficiently reliable.

APPENDIX I

To evaluate the integral

$$\int_0^T\left(1-\frac{t}{T}\right)dt[1-(1-e^{-\beta t})^{\frac{1}{2}}]dt$$

we note that we neglect terms smaller than $\beta^{-1}e^{-\beta T}$ if we extend the integration to infinity, since

$$1-(1-e^{-\beta T})^{\frac{1}{2}}\leq e^{-\beta t}$$

for real, non-negative t and β. We then have

$$I=\int_0^\infty\left(1-\frac{t}{T}\right)[1-(1-e^{-\beta t})^{\frac{1}{2}}]dt$$

$$=\left\{\left(1+\frac{1}{T}\frac{\partial}{\partial\lambda}\right)\int_0^\infty e^{-\lambda t}[1-(1-e^{-\beta t})^{\frac{1}{2}}]dt\right\}_{\lambda=0}$$

$$=\left\{\left(1+\frac{1}{T}\frac{\partial}{\partial\lambda}\right)\left[\frac{1}{\lambda}-\frac{1}{\beta}\int_0^1 x^{(\lambda/\beta)-1}(1-x)^{\frac{1}{2}}dx\right]\right\}_{\lambda=0}$$

$$=\left\{\left(1+\frac{1}{T}\frac{\partial}{\partial\lambda}\right)\left[\frac{1}{\lambda}-\frac{1}{\beta}\left(\frac{\Gamma(\lambda/\beta)\Gamma(3/2)}{\Gamma(\lambda/\beta+3/2)}\right)\right]\right\}_{\lambda=0}$$

$$=\left\{\beta^{-1}\left[1+(\beta T)^{-1}\frac{\partial}{\partial z}\right]\left[z^{-1}-\left(\frac{\Gamma(z)\Gamma(3/2)}{(z+\frac{1}{2})\Gamma(z+\frac{1}{2})}\right)\right]\right\}_{z=0}.$$

From

$$\frac{d}{dz}lg\frac{\Gamma(z+\frac{1}{2})}{\Gamma(z)}=2\sum_{k=0}^{\infty}\frac{(-)^k}{2z+k}$$

one obtains†

$$\Gamma(z)/\Gamma(z+\tfrac{1}{2})=\pi^{-\frac{1}{2}}z^{-1}\exp\left\{-\sum_{n=1}^{\infty}\frac{\kappa_n}{n!}(iz)^n\right.$$

with

$$\kappa_1=2i\,lg2$$

and

$$\kappa_n=(n-1)!(2i)^n(1-2^{1-n})\zeta(n)\quad\text{for}\quad n\geq 2,$$

where $\zeta(n)$ is the Riemann zeta-function.‡ We thus have

$$\frac{\Gamma(z)}{\Gamma(z+\frac{1}{2})}=\pi^{-\frac{1}{2}}2^{-1}\left(1-i\kappa_1 z+z^2\frac{\kappa_2-\kappa_1^2}{2}+\cdots\right)$$

and

$$z^{-1}-\frac{\Gamma(3/2)\Gamma(z)}{(z+\frac{1}{2})\Gamma(z+\frac{1}{2})}=z^{-1}\{1-(1-2z+4z^2+\cdots)$$

$$\times(1-i\kappa_1 z+z^2\tfrac{1}{2}(\kappa_2-\kappa_1^2)+\cdots)\}$$

$$=(2+i\kappa_1)-z(4+\tfrac{1}{2}(\kappa_2-\kappa_1^2)+2i\kappa_1)$$

and

$$I=\beta^{-1}\{(2+i\kappa_1)-(\beta T)^{-1}[4+\tfrac{1}{2}(\kappa_2-\kappa_1^2)+2i\kappa_1]\}$$

$$=\beta^{-1}\left\{2(1-lg2)-(\beta T)^{-1}\left[4(1-lg2)-\frac{\pi^2}{6}+2(lg2)^2\right]\right\}$$

$$=\beta^{-1}\{0.61-0.54(\beta T)^{-1}\}.$$

APPENDIX II

Outside of the range of the sample, defined by $\vartheta(a,T)=1$ for all $a\leq l$ and $\vartheta(a,T)=0$ for all $a\geq u$, the

† For detailed derivation, see A. J. F. Siegert, "On the roots of Markoffian random functions," Rand Corporation Report RM-447.

‡ See, e.g., Jahnke Emde, Tables of Functions, p. 269 ff.

integral (3.17) can be obtained analytically as a series in powers of $(-\sigma/l)$ and (σ/u). We have

$$\delta_2^* \equiv \int_l^u [\vartheta(a,T)-\phi(a)]^2 \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da + \int_u^\infty \phi^2(a) \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da + \int_{-\infty}^l [1-\phi(a)]^2 \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da,$$

where

$$\phi(a) \equiv \frac{1}{(2\pi\sigma^2)^{\frac{1}{2}}} \int_a^\infty \exp\left(-\frac{y^2}{2\sigma^2}\right) dy = \frac{1}{(\pi)^{\frac{1}{2}}} \int_{a/\sigma\sqrt{2}}^\infty \exp(-x^2) dx = \tfrac{1}{2}\left[1 - \frac{2}{(\pi)^{\frac{1}{2}}} \int_0^{a/\sigma\sqrt{2}} \exp(-x^2) dx\right] \cong \frac{1}{(2\pi)^{\frac{1}{2}}} \frac{\sigma}{a} \exp\left(-\frac{a^2}{2\sigma^2}\right)$$

valid for $a \gg \sigma$, and

$$I_u \equiv \int_u^\infty \phi^2(a) \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da \simeq \sigma(2\pi)^{-\frac{1}{2}} \int_u^\infty \frac{da}{a^2} \exp\left(-\frac{a^2}{2\sigma^2}\right),$$

which by integration by parts and use of the asymptotic series for the error function yields

$$I_u \simeq (2\pi)^{-\frac{1}{2}} (\sigma/u)^3 \exp\left(-\frac{u^2}{2\sigma^2}\right) \simeq \frac{1}{2\pi}\left(\frac{\sigma}{u}\right)^2 \phi(u)$$

valid for $u \gg \sigma$ and

$$I_l \equiv \int_{-\infty}^l [1-\phi(a)]^2 \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} = \int_{-l}^\infty \phi^2(a) \frac{\exp(a^2/2\sigma^2)}{(2\pi\sigma^2)^{\frac{1}{2}}} da \simeq \frac{1}{2\pi}\left(\frac{\sigma}{l}\right)^2 \phi(-l)$$

valid for $-l \gg \sigma$.

APPENDIX III

To evaluate the integral

$$\int_0^T (1-t/T)\, dt\, lg(1-\epsilon^{-\beta t}),$$

we note that the integral can be extended to ∞ incurring an error of order $\epsilon^{-\beta T}$ only. We then need—for $s=3$ and $s=2$—the integral

$$\int_0^\infty t^{s-2} dt lg(1-\epsilon^{-\beta t}) = \left[\frac{t^{s-1}}{s-1} lg(1-\epsilon^{-\beta t})\right]_0^\infty = -\frac{1}{s-1} \int_0^\infty t^{s-1} \beta dt \epsilon^{-\beta t} (1-\epsilon^{-\beta t})^{-1} = -\frac{\beta^{1-s}}{s-1} \int_0^\infty x^{s-1} dx \epsilon^{-x} (1-\epsilon^{-x})^{-1}.$$

The last integral is equal to $\Gamma(s)\zeta(s)$,§ so that

$$\int_0^\infty \left(1-\frac{t}{T}\right) dt\, lg(1-\epsilon^{-\beta t}) = -\beta^{-1}\zeta(2) + \beta^{-2}T^{-1}\zeta(3).$$

§ E. T. Whittaker and G. N. Watson, *Modern Analysis* (Cambridge University Press, London, 1935), p. 266.

The Admittance of a Diode with a Retarding Field

J. J. FREEMAN
National Bureau of Standards, Washington, District of Columbia
(Received February 4, 1952)

The contribution of space charge to the admittance of a diode with a retarding field is calculated by computing the instantaneous electronic current induced in an external circuit. A discussion of the variation of the susceptance of a typical diode with frequency is given.

I. INTRODUCTION

THE contribution to a diode's input conductance induced by transit time and space charge effects was computed by Begovich[1] for the case of a diode with a retarding field. His method involved energy considerations which necessarily limited the investigation to a consideration of the conductance alone. In this paper both the conductance and admittance of such a diode are computed simultaneously by evaluating the instantaneous electronic current induced in a circuit connected externally to the diode. The expression for the conductance is found to be the same as that derived by Begovich. The susceptance is found to be positive (capacitative) at low frequencies and negative (inductive) at high frequencies, its variation with frequency roughly resembling the susceptance of a resistance, capacitance, and inductance in series.

II. THEORY*

In the following it is assumed that the plate voltage will be more than 1 volt below that of the cathode, and that the cathode-anode spacing d will be sufficiently small so that the potential distribution between the cathode and anode may be assumed to be linear. It is further assumed that the cathode and anode consist of two parallel planes sufficiently close so that edge effects may be neglected.

Consider an electron at distance x from the cathode (Fig. 1) traveling with velocity $\dot{x}$. Regardless of its interaction with other electrons, an electron will induce a current[2]

$$I(t)=-e\dot{x}/d$$

in the external circuit, where $\dot{x}$ is the velocity of the electron. The total induced current is simply the sum over the electron velocity distribution of the elementary induced currents, and the immediate problem is to calculate the velocity of a typical electron. The effect of electron interactions or collisions is obtained by solving Poisson's equation for the average potential distribution, since each electron experiences the resultant of a large number of long-range Coulomb interaction forces. In the present case the potential distribution $V(x)$ may be assumed to be linear,† and the ac variation of potential is assumed small compared to its dc value.

Let V_1 be the dc (negative) potential on the anode, $V_2\cos(\omega t+\phi)$ the ac potential, and let $\alpha=V_2/V_1$. We shall make the usual small-signal theory approximation that $\alpha\ll 1$.

Then

$$V(x)=(V_1x/d)[1+\alpha\cos(\omega t+\phi)]. \tag{1}$$

The trajectory of each electron is determined by the solution of

$$\ddot{x}=-\beta[1+\alpha\cos(\omega t+\phi)], \tag{2}$$

where

$$\beta=eV_1/md$$

and

$$e=-1.59\times 10^{-19}\text{ coulombs}.$$

The first integral of (2) yields

$$\dot{x}=v_0-\beta\{t-t_0+2\alpha/\omega\sin[\omega(t-t_0)/2]\times\cos[\omega t_0+\phi+\omega(t-t_0)/2]\}, \tag{3}$$

where v_0 is the initial velocity of the electron emitted from the cathode ($x=0$) at time $t=t_0$. The velocity at any instant t depends only on the time elapsed since emission $t-t_0$, the initial velocity, and the phase of the generator at the time of emission, $\omega t_0+\phi$.

The position of the electron at time t is

$$x=v_0(t-t_0)-\beta\left\{\frac{(t-t_0)^2}{2}-\frac{\alpha}{\omega^2}[\cos(\omega t+\phi)-\cos(\omega t_0+\phi)+\omega(t-t_0)\sin(\omega t_0+\phi)]\right\}. \tag{4}$$

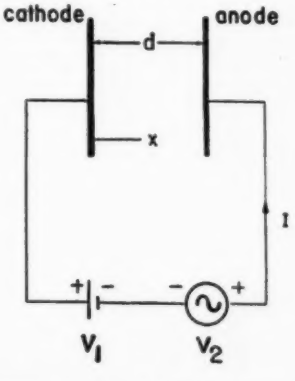


FIG. 1. Schematic of idealized diode.

[1] N. Begovich, J. Appl. Phys. **20**, 457–461 (1949).
* MKS units are used unless otherwise specified.
[2] W. Shockley, J. Appl. Phys. **9**, 635–636 (1938); C. Bakker and G. de Vries, Physica **1**, 1045 (1934).

† For a justification of this assumption see reference 1.

The transit time τ is that value of $t-t_0$ (other than zero) which satisfies Eq. (4) for $x=0$. Let $\tau=\tau_0+\alpha\tau_1$ as a first-order approximation in α, and one finds

$$\tau_0=2v_0/\beta$$

and

$$\tau_1=\frac{2}{\tau_0}\left[-\frac{2}{\omega^2}\sin\left(\omega\frac{\tau_0}{2}+\phi\right)\sin\frac{\omega\tau_0}{2}+\frac{\tau_0}{\omega}\sin(\omega t_0+\phi)\right]. \quad (5)$$

Let N be the number of electrons emitted per second. Of these,

$$f(v_0)dv_0=\frac{mv_0}{kT}\exp-\frac{mv_0^2}{2kT}dv_0 \quad (6)$$

is the fraction having initial velocity between v_0 and v_0+dv_0. Then the induced current at time t_0, $\delta I(t)$, owing to electrons with initial velocity v_0, is obtained by summing the contributions of this velocity class emitted during the time interval $t-\tau$ to t.

$$\delta I(t)=-\frac{eN}{d}f(v_0)\int_{t-\tau}^{t}\dot{x}(t,t_0)dt_0. \quad (7)$$

The total induced current is

$$I(t)=-\frac{Ne}{d}\int_0^{v_m}f(v_0)\int_{t-\tau}^{t}\dot{x}(t,t_0)dt_0dv_0, \quad (8)$$

where $v_m\simeq(2eV_1/m)^{\frac{1}{2}}$ is the initial velocity required to reach the plate. It will be seen later that the precise value of v_m is unimportant. Neglecting quantities involving α^2 or higher, and letting

$$\delta=\omega\tau_0/2,$$

$$\int_{t-\tau}^{t}\dot{x}(t,t_0)dt_0=\frac{4\alpha\beta}{\omega^2}\sin(\omega t+\phi-\delta)[\sin\delta-\delta\cos\delta] \quad (9)$$

$$I(t)=-\frac{4Nem\alpha\beta}{dkT}\int_0^{v_m}\left[v_0\exp\left(-\frac{mv_0^2}{2kT}\right)\Big/\omega^2\right]\times[\sin\delta-\delta\cos\delta]\times[\sin(\omega t+\phi)\cos\delta-\sin\delta\cos(\omega t+\phi)]dv_0. \quad (10)$$

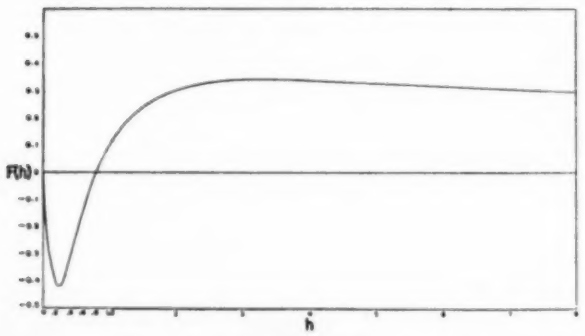


FIG. 2. Graph of $F(h)$.

Let

$$A=-\frac{4\alpha\beta Nme}{\omega^2dkT}\int_0^{v_m}v_0\exp\left(-\frac{mv_0^2}{2kT}\right)\times[\sin\delta-\delta\cos\delta]\cos\delta dv_0, \quad (11)$$

$$C=\frac{4\alpha\beta}{\omega^2}\frac{Nme}{dkT}\int_0^{v_m}v_0\exp\left(-\frac{mv_0^2}{2kT}\right)\times[\sin\delta-\delta\cos\delta]\sin\delta dv_0, \quad (12)$$

so that

$$I(t)=A\sin(\omega t+\phi)+C\cos(\omega t+\phi).$$

Then if G is the conductance and B is the susceptance, it follows that

$$G=\frac{4Nme\beta}{dkT\omega^2V_1}\int_0^{v_m}v_0\exp\left(-\frac{mv_0^2}{2kT}\right)\times[\sin^2\delta-\delta\sin\delta\cos\delta]dv_0, \quad (13)$$

$$B=\frac{4Nme\beta}{dkT\omega^2V_1}\int_0^{v_m}v_0\exp\left(-\frac{mv_0^2}{2kT}\right)\times[\sin\delta-\delta\cos\delta]\cos\delta dv_0. \quad (14)$$

Changing the variable of integration from v_0 to δ, and letting

$$h=m\beta^2/2kT\omega^2, \quad (15)$$

$$\delta_0=\omega[2d/\beta]^{\frac{1}{2}}, \quad (16)$$

$$I_0=Ne, \quad (17)$$

one finally obtains, after integrating by parts,

$$G=\frac{4I_0kT}{eV_1^2}\Big[\exp(-h\delta_0^2)\times\{(1+h)\cos2\delta_0+h(\delta_0\sin2\delta_0-1\}+\frac{(2+h)}{h}\int_0^{\delta_0}\exp(-h\delta^2)\sin2\delta d\delta-1\Big], \quad (18)$$

$$B=\frac{-4I_0kTh}{eV_1^2}\Big[\exp(-h\delta_0^2)\times\Big[\sin2\delta_0-\delta_0\cos2\delta_0-\delta_0+\frac{1}{h}\sin2\delta_0\Big]-\Big(1+\frac{2}{h}\Big)\int_0^{\delta_0}\exp(-h\delta^2)\cos2\delta d\delta+\int_0^{\delta_0}\exp(-h\delta^2)d\delta\Big]. \quad (19)$$

Since $h\delta_0^2=eV_1/kT$ will always exceed 10 for the

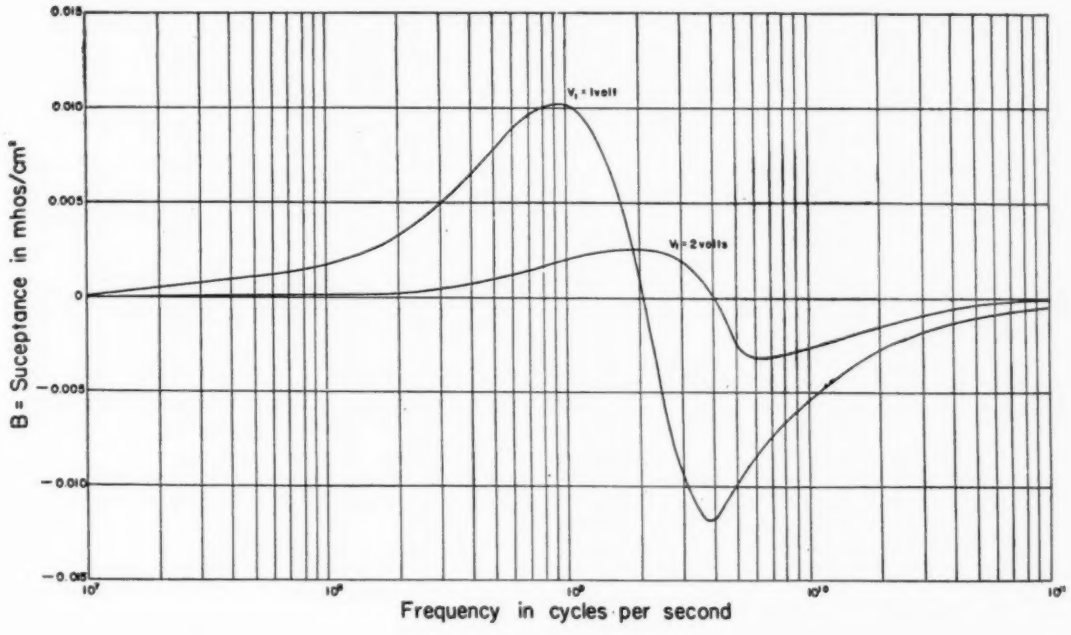


FIG. 3. Susceptance of diode.

voltages under consideration, namely $|V_1|>1$ volt, and since the integrands in the integral expressions above are negligible for $\delta=\delta_0$, the upper limit of integration may be taken as infinity.

Equation (18) agrees with the expression for the conductance derived by Begovich and will not be considered further. Taking the upper limit of the integral in Eq. (19) as infinity, and neglecting the expression in square brackets which has the factor $\exp(-h\delta_0^2)$, one finally obtains

$$B=\frac{2I_0kT(\pi)^{\frac{1}{2}}}{eV_1^2}F(h), \tag{20}$$

where

$$F(h)=(h)^{\frac{1}{2}}\left[\left(1+\frac{2}{h}\right)\exp\left(-\frac{1}{h}\right)-1\right]. \tag{21}$$

From Fig. 2, it may be seen that $F(h)$ has zeros at $h=0$, $h=0.8$ and $h=\infty$. For h small,

$$F(h)=-(h)^{\frac{1}{2}}, \tag{22}$$

so that it follows from Eqs. (20) and (15), that for ω large,

$$B=-\frac{I_0}{\omega dV_1}\frac{2\pi kT}{m}, \tag{23}$$

which is the susceptance of an equivalent inductance L, where

$$L=\frac{V_1d}{I_0}\left(\frac{m}{2\pi kT}\right)^{\frac{1}{2}}. \tag{24}$$

For h large (low frequencies) $F(h)=1/h^{\frac{1}{2}}$, and Eq. (20) becomes

$$B=\frac{2\omega I_0d}{e^2V_1^3}(kT)^{\frac{1}{2}}(2\pi m)^{\frac{1}{2}}, \tag{25}$$

which is the susceptance of an equivalent capacitance C, where

$$C=\frac{2I_0(kT)^{\frac{1}{2}}d(2\pi m)^{\frac{1}{2}}}{e^2V_1^3}. \tag{26}$$

III. EXAMPLE

The susceptance of a diode is plotted in Fig. 3 using the constants in the example cited by Begovich, namely, $d=8.89\times10^{-3}$ cm, $T=10^3$, $I_0=0.1$ amps/cm². For $V_1=1$ volt, the low frequency susceptance due to space charge equals that of a capacitor of 2.7 $\mu\mu$f; the susceptance increases with frequency until it reaches a maximum at 950 Mc, decreases to zero, reaching a minimum, and then approaches zero asymptotically behaving like a shunt inductance of 29 $\mu\mu$h.

The susceptance of the same tube for $V_1=2$ volts is also plotted in Fig. 3, the values being derived from the first curve by utilizing the similitude relation

$$B(fV_1/V_1', V_1')=B(f, V_1)[V_1/V_1']^2,$$

which is readily determined from Eqs. (20) and (15).

The Creep of Zinc Single Crystals*

L. SLIFKIN† AND W. KAUZMANN
Frick Chemical Laboratory, Princeton University, Princeton, New Jersey
(Received February 22, 1952)

The transient creep of single crystals prepared from 99.999 percent zinc was studied at 35°C and at strain rates of the order of 10^{-6}/min. The results are given by the relation: creep rate$=at^{-n}$, where a and n are constants and t is the time after applying the load. This is of the form expected from Andrade's equation, but the exponent n had only approximately the value $\frac{2}{3}$ found by Andrade and varied from experiment to experiment. An additional instantaneous plastic strain occurs on applying the stress at the start of a creep test. It was found to be proportional to the subsequent transient creep rate (as measured by a in the above equation) in a series of tests made on a given crystal. In addition to the process of recovery from strain-hardening whereby a zinc crystal completely resoftens in a few days at 35°C, there appears to be a rapid, but very limited, stage of recovery, the effects of which quickly disappear upon further deformation. Some crystals exhibited a decrease in plasticity when allowed to rest for about two days at 35°C after a creep test. The development of this "rest-hardening" is dependent to some extent on the occurrence of creep prior to the rest period. Zinc crystals were observed to contract with time after removing the load in a creep test. The extent of this contraction varied among specimens and among different tests made with the same specimen and was of the order of 10^{-6} cm/cm or less. This appears to be the first report of such an after effect in single crystals of a pure metal. The results are discussed in relation to the findings of other workers and are interpreted in terms of recent models of the metallic state.

INTRODUCTION

IF a sufficiently large load is applied to a metal crystal, one finds that in addition to the instantaneous elastic strain a further deformation occurs which increases slowly with time.[1] This time-dependent flow under constant load is known as creep. How great the "sufficiently large" stress must be depends on many factors, such as the sensitivity of the strain gauge, the temperature, and the condition of the specimen. We shall here be concerned with axial tensions applied to cylindrical single crystals of zinc. The use of single crystals eliminates the effect of grain boundaries on the creep properties and should greatly simplify the interpretation of the results. Under these conditions elongation proceeds by slip, a process involving the shearing of certain crystallographic planes across one another, much as though the crystal were a stack of playing cards. Thus, the rate of strain depends not on the applied tensile stress, but on the component of this stress which acts to produce shear in a specific direction in the slip plane.

Andrade[2] found that under a variety of conditions, the results of a creep test could often be represented by an equation which may be written[3]

$$\gamma=\gamma_0+\beta t^{\frac{1}{3}}+Kt, \qquad \text{(I)}$$

where γ is the shear strain at time t after applying the stress, γ_0 is an instantaneous plastic strain, and β and K are constants. The term in Andrade's equation containing β represents a flow called *transient creep*, the rate of which decreases with time. The term containing K represents a steady-state flow, called *quasi-viscous creep*. A third type of creep, in which the strain rate increases with time, is sometimes observed in the later stages of a test; the import of this process is not clear, and it is still maintained by some to be an artifact.[4]

Little is known of the dependence of γ_0, β, and K on such variables as stress and temperature or of how their values compare for single crystals and polycrystals. It appears that β/K is much larger than 1 min$^{\frac{1}{3}}$ at low temperatures and low stresses, but that the ratio decreases as these variables increase. The dependence of the parameter K on the temperature and the stress has been discussed elsewhere.[5]

The primary purpose of this work was to study the transient creep of zinc single crystals at 35°C in the range of shear rates 10^{-6} to 10^{-5}/min over intervals of a few hundred minutes. Such tests do little permanent damage to the crystal, since it was found that annealing usually restored the plastic properties of the specimen to their initial condition. Thus it was possible to do many experiments on the same specimen, thereby eliminating the effect of the variability found to occur among different crystals.

EXPERIMENTAL

Preparation and Treatment of the Single Crystals

The crystals were grown by the method of Bridgman,[6] using 99.999 percent zinc generously given to us by

* Taken from a thesis submitted by L. Slifkin in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Princeton University.

† Present address: Department of Physics, University of Illinois, Urbana, Illinois.

[1] E. Schmid and W. Boas, *Plasticity of Crystals* (F. A. Hughes and Company, London, 1950). F. Seitz, *Physics of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).

[2] E. N. da C. Andrade, Proc. Roy. Soc. (London) A84, 1 (1910).

[3] A. H. Cottrell and V. Aytekin, Nature 160, 328 (1947).

[4] See, for example, A. H. Sully, *Metallic Creep* (Butterworth Scientific Publications, London, 1949).

[5] W. Kauzmann, Trans. Am. Inst. Met. Engrs. 143, 57 (1941).

[6] P. W. Bridgman, Proc. Am. Acad. Arts Sci. 58, 165 (1923); 60, 305 (1925).

the New Jersey Zinc Company. A few specimens were grown in graphite-lined glass tubes, but the great majority were prepared in the following manner. The metal was first melted in a Vycor flask connected to a vacuum system. When the pressure had fallen to 10^{-5} mm of Hg the zinc was considered to have been sufficiently degassed. The flask was removed from the vacuum system and the metal cast into rods by drawing it up into graphite-lined Pyrex tubes. Each rod was placed in a crucible made by drilling an axial $\frac{1}{4}$-inch hole in a cylinder of spectroscopically pure graphite $\frac{1}{2}$ inch in diameter. The zinc was melted into the crucible under vacuum to preclude air holes in the final product. The crucible and its charge were removed from the vacuum system and then lowered at a rate of 5 cm/hr through a furnace having a sharp vertical temperature gradient. Nitrogen, previously passed through alkaline pyrogallol solution and magnesium perchlorate drying agent, was slowly led in through the bottom of the furnace. In the preparation of some later crystals, the crucible was sealed in a glass tube filled with argon purified over hot calcium. Here the lowering rate was 3.3 cm/hr.

The orientations of the crystals were determined optically, using the cork ball method of Bridgman.[7] The etchant employed was the acidic $CuCl_2$ solution recommended by Barrett.[8] This solution has the disadvantage of leaving dark deposits of copper or copper oxide on the surface, but it produces well-defined reflections. It is believed that the error in the measurement of orientations was about 2°. It is interesting that for half of the crystals produced, the angle between the c axis and the specimen axis was in the range 64°-79°.

Before each run, the crystal was annealed for three hours at 210°C. In the first experiments the annealing was conducted in air, but in most of the work an atmosphere of commercial nitrogen was used.

Prior to mounting the extensometers on a crystal, the pointed end of the specimen was carefully cut off with snippers. This operation would be expected to cause local lattice bending, some of which might be transmitted through the crystal in the c direction.[9] Since the crystals were only of $\frac{1}{4}$-inch diameter and the c axis was always inclined to the specimen axis, this transmitted deformation should have affected only the extreme ends, which were not included in the gauge length.

Two sleeves of glass tubing with roughened surfaces were next placed end to end over the specimen, and the outside end of each fixed to the zinc, using calcium carbonate—water glass cement. The distance between the two cemented ends represented the gauge length of the sample, which was known to 0.05 cm. Chucks were then carefully attached to each end of the crystal. The effects on the crystal lattice of tightening the chucks were probably significant only in the regions near the ends of the crystal, which were not included in the gauge length.

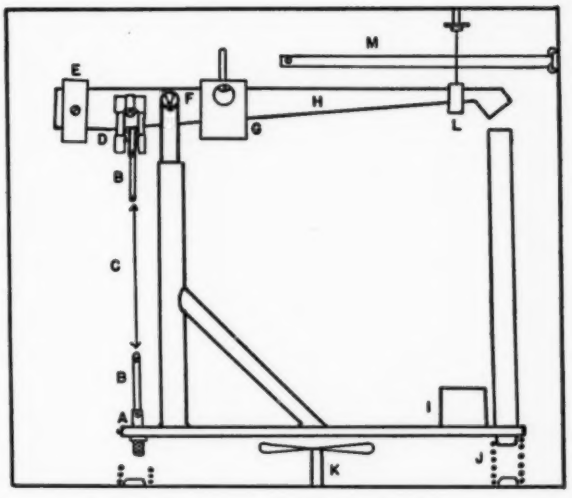


FIG. 1. Creep testing frame. The crystal, with extensometers, is mounted at C and connected to the anchor piece A (height adjustable) and the beam H by means of nickel strips B. The beam is pivoted on the knife-edge F and is counterbalanced by the weight E. The stress is produced by the weight G and is applied axially through two perpendicular knife-edges D. The stress can be released by raising the beam stirrup L, which is controlled from outside the enclosing box. M is a hook for moving the weight G and is also externally controlled. The entire assembly is supported by springs J to eliminate vibrations, and maintained level by ballast I. A fan K circulates the air in the system.

Measurement of the Creep

The extensometers were similar to the modified Martens extensometer[4] and operated on the optical lever principle. To minimize errors due to bending of the specimen while under tension, two extensometers were mounted on opposite sides of the crystal. They were suspended by springs from a rack held by the upper chuck so that the crystal did not support their weight, and were then firmly fixed to the glass sleeves by means of springs.

In a creep test, the crystal was subjected to tensional stress. The device for applying this tension is shown in Fig. 1.‡ Not shown in the diagram is a spring just below "D" which should soften the shock resulting from application of the load upon lowering the beam.

This device was contained in an insulated box, in which the temperature was maintained at 35.25°C by a thermoregulator comprised of a thermistor, Wheatstone bridge, galvanometer, and a photocell circuit. The temperature of the specimen was estimated with a thermocouple, one junction of which was inserted into

[7] A. T. Gwathmey, J. Phys. Chem. **44**, 35 (1940).
[8] C. S. Barrett, *Structure of Metals* (McGraw-Hill Book Company, Inc., New York, 1943).
[9] D. C. Jillson, Trans. Am. Inst. Met. Engrs. (J. Metals) **188**, 1009 (1950).

‡ We wish to thank Dr. Nadai of the Westinghouse Laboratories for advice in the design of this apparatus. The apparatus was constructed in the shops of the Westinghouse Laboratories.

TABLE I. Values of n in the transient creep equation. For runs in which quasi-viscous creep was appreciable, the slopes of the initial, linear parts of the log creep rate *vs* log time plots are used. "Run $x-y$" means the yth creep test performed on crystal No. x. Most of the values are good to better than ± 0.05, the rest to ± 0.10.

Run	n	Run	n	Run	n
5– 1	0.71	13– 7	0.75	19– 4	0.75
8– 1	0.73	8	0.73	5	0.84
3	0.86	9	0.73	6	0.83
4	0.68	10	0.74	7	0.83
5	0.86	11	0.74	8	0.67
6	0.70	12	0.71	9	0.81
7	0.74	13	0.73	10	0.82
8	0.72	14	0.77	12	0.72
9	0.68	15	0.72	20– 1	0.70
9– 1	0.68	16	0.62	3	0.70
5	0.68	17	0.69	22– 1	0.69
6	1.28	18	0.70	24– 1	0.60
11– 1	0.75	19	0.73	28– 1	0.93
2	0.67	20[a]	0.54	2[a]	0.82
4	0.69	21	0.61	29– 1	0.83
5	0.68	22	0.56	2	0.85
6	0.75	17– 5	1.00	3	0.64
7	0.69	7	0.69	30– 1	0.69
13– 1	0.85	8	0.64	2	0.61
2	0.85	9	0.64	31– 1	0.75
3	0.85	12	0.51	2	0.85
4	0.84	19– 1	0.82	3	0.81
5	0.82	2	0.82	4	0.92
6	0.75	3	0.80	5	0.80
				6	0.72

[a] The specimen was accidentally bent and then straightened before this run.

a near-by dummy crystal. Corrections due to thermal expansion were taken into account in computing the strain. They were almost always small compared to the plastic elongations.

With this apparatus, each reading of length could be estimated to the nearest 2.5×10^{-6} cm, with an error of less than twice this amount. With the usual gauge length of 10 cm, this gives a maximum error in the strain of 5×10^{-7}. Since each measurement of elongation involved taking the difference between two readings, we may estimate a maximum error of about 10^{-6} in the determination of the strain, or about 100 A per cm of gauge length.

With a 10-cm gauge length, creep rates as low as about 0.3×10^{-5}/min could be measured if readings of elongation were made once a minute. Rates an order of magnitude lower than this could be measured if the rate was changing slowly enough with time to permit a longer wait between readings. For still lower creep rates, the necessary time intervals involved were such that corrections due to expansion or contraction of the specimen resulting from temperature changes were often appreciable compared to the small plastic elongations involved.

The apparatus was tested by measuring the coefficient of thermal expansion of a copper rod and the Young's modulus of a steel rod. The resulting plots of length *versus* temperature and length *versus* stress, respectively, were good straight lines with little scatter of the points, and the slopes gave values of the constants in agreement with the accepted values.

RESULTS

In this investigation, approximately 80 creep tests were performed on 19 single crystals. The specimens were annealed as described above after each test; if the crystals were not made to creep at too high a rate, this treatment returned them to the same state, as indicated by the creep behavior. The reproducibility obtained in a series of tests made with the same crystal is discussed below. Individual creep tests lasted from 30 to 500 minutes. Although the significant flow in these experiments is a shear, the results will often be discussed in terms of tensile strain, which for a given zinc crystal is related to the shear by the equation

$$\epsilon = s \sin\alpha \cos\beta,$$

where ϵ is the tensile strain due to shear in the slip direction, s is the shear strain in the slip direction, α is the angle between the c axis and the specimen axis, and β is the angle between the projection of the specimen axis on the basal plane and the nearest slip direction. The angle β can never be greater than 30° for zinc, and in most of the specimens used in this work α was between 64° and 79°. Thus, the shear rate was usually larger than the elongation rate by only ten or twenty percent.

General Behavior

Most of the creep tests involved only transient creep and fitted well the equation

$$\text{Creep rate} = at^{-n}. \qquad \text{(II)}$$

Here t is the time after application of the stress and a and n are parameters which are constant for a given run. We shall refer to this relation as the "transient creep equation" and to a as the "transient creep parameter." One sees that Eq. (II) may be derived from Andrade's law (Eq. (I)) if $K=0$ (i.e., if there is no quasi-viscous

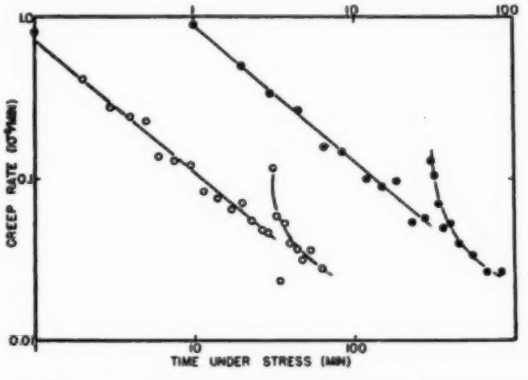


FIG. 2. Effect on creep behavior of one-minute (left-hand curve) and 10-minute (right-hand curve) rest after 30 minutes of creep. The time scale on the bottom refers to the curve on the left; that on top to the curve on the right. Resolved shear stress $=2.41\times10^8$ dynes/cm^2.

creep) and if our n is equal to $\frac{2}{3}$, whereupon $a=\beta/3$. On plotting the results as the logarithm of the creep rate against the logarithm of the time, however, the straight line obtained for a given creep test seldom had a slope of exactly $-\frac{2}{3}$. The values obtained for n are presented in Table I. One sees that there is no strikingly closer agreement among the values of n for runs on the same crystal than among those of different crystals. No dependence of n on the orientation of the specimen was apparent.

For runs in which a was larger than 5 or 10×10^{-5} $\text{min}^{(n-1)}$, quasi-viscous creep was often quite prominent. In some cases, this steady-state creep led to an accelerating creep. The creep rate did not always increase monotonically with time after the onset of accelerating creep, as was also observed by Miller.[10]

It should be pointed out that the creep of zinc crystals does not proceed uniformly with time but that the rate fluctuates considerably during a run.[11] In Figs. 2 and 3, for example, it is seen that there is an appreciable scatter of the points about any smooth curve drawn through them. This scatter is not due to experimental error but represents the true behavior of the flowing crystal. It is interesting to note that some crystals consistently creep more uniformly than others i.e., coarseness of fluctuations appears to be a characteristic of a crystal.

Since all creep tests were performed at 35.25°C, nothing was learned of the temperature dependence of transient creep. Some information concerning the stress dependence was obtained, however, by performing a series of tests at different stresses on a given crystal, annealing for 3 hours at 210°C before each run. It was found, as Tyndall[12] has reported, that the transient creep rate increased exponentially with stress:

$$a=(\text{const})\ \exp(b\sigma),$$

where a is the transient creep parameter, σ is the resolved shear stress, and b had the values 2.8 and 5.5×10^{-6} cm²/dyne for the two crystals studied. The crystal giving the value 2.8×10^{-6} contained impurities and had a considerably greater tensile strength than the usual pure specimens. It is questionable, however, how much significance can be attached to the lower value of b found for it, since these values may only be valid to within a factor of about 2. This is because a could be varied by little more than a factor of 10; for very small stresses the creep rate was too small to be measured, while for large stresses quasi-viscous creep became prominent.

Satisfactory reproducibility in the results of a number of creep tests made on any given crystal was obtained. The values of the transient creep parameter a in many tests with a particular crystal at a given stress generally agreed within a range of a factor of 2. This is not an unacceptably large variation when one considers that it is the logarithm of the rate that is proportional to such factors as stress. Successive runs on a crystal showed no trends in the variation of a or n. It was therefore concluded that the annealing treatment after each run returned the crystal to approximately the same state. This conclusion may be valid only for the case of transient creep. It was often observed that after extensive quasi-viscous creep, the creep behavior in succeeding runs gradually changed, with the crystal usually becoming softer.

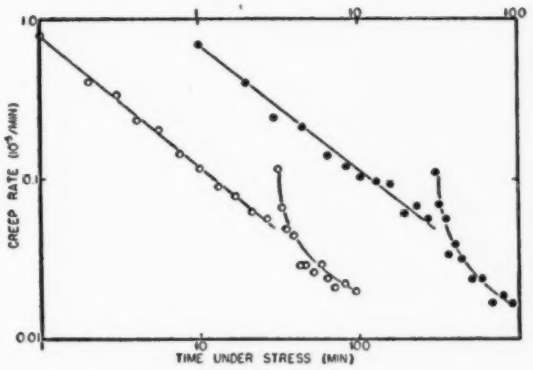


FIG. 3. Same as Fig. 2 except 50- and 250-minute rest periods instead of one and ten minutes, respectively.

The results of tests on different crystals were not as reproducible as those of tests on the same crystals, and a large variation in tensile strength from specimen to specimen was found. Since the shear rate = creep rate ÷ $\sin\alpha\cos\beta$, creep tests on different crystals should be comparable when the values of $a/\sin\alpha\cos\beta$ are equal, where a is the transient creep parameter. The resolved shear stress required to produce a value of $a/\sin\alpha\cos\beta$ equal to 10^{-5} $\text{min}^{(n-1)}$ lay in the range $2.0-2.5\times10^{6}$ dynes/cm² for seven crystals grown in nitrogen and two grown in argon. For four crystals grown in argon the stress required was $1.1-1.4\times10^{6}$ dynes/cm². Because of the exponential dependence of shear rate on stress, one could at best predict only the order of magnitude of the rate to be expected upon applying a given stress to a new crystal. It is not known whether the softness of four of the argon-grown specimens was a result of the substitution of argon for nitrogen, or whether it was due to other factors involved in changing to the different technique employed in growing crystals in argon. Nor is it known why only four of the six argon-grown crystals were soft.

It was found that if the crystals were annealed in air they often became harder in later creep tests. In one instance an overnight anneal in air hardened the specimen to such an extent that a shear stress of 2.52×10^{6} dynes/cm² produced a somewhat lower creep rate than 2.41×10^{6} dynes/cm² had before the anneal. After a light etching in dilute HCl and a brief further anneal, the creep rate produced by the higher stress was greater by a factor of three. It thus appears that

[10] R. F. Miller, Trans. Am. Inst. Met. Engrs. **122**, 176 (1936).
[11] E. Schmid and M. A. Valouch, Z. Physik **75**, 531 (1932).
[12] E. P. T. Tyndall, J. Appl. Phys. **21**, 939 (1950).

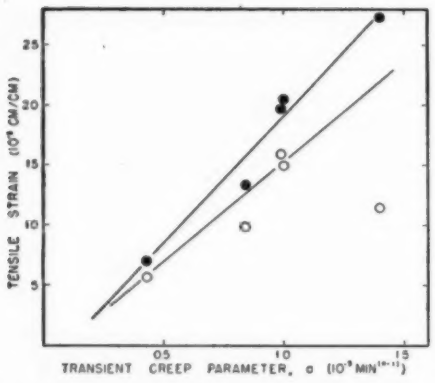


FIG. 4. Proportionality between the instantaneous initial plastic elongation and the transient creep parameter.

the hardening can be attributed to the effect of an oxide coat, as reported by Harper and Cottrell.[13]

No slip bands were observed on the specimens, even after many creep tests with the same crystal. The crystals were not, however, examined under magnification.

Instantaneous Initial Plastic Strain

The equation of Andrade, giving the strain as a function of time, contains a term γ_0 which represents a plastic strain occurring "instantaneously" upon application of the stress. This initial plastic elongation was measured in several creep tests made on each of three crystals. It was found that in a series of runs, all made on the same crystal, this initial extension was proportional to the transient creep parameter α (which in turn is approximately equal to $\beta/3$). An example of the experimental results is shown in Fig. 4. One may therefore replace the γ_0 of Andrade's equation by $\delta\beta$, where δ is a constant characteristic of the creep behavior of a given crystal in a series of tests. Values of δ obtained for the three crystals were 1.6, 10.2, and 5.5 min$^{\frac{1}{2}}$, all $\pm\sim 0.5$.

Recovery at 35°C

Experiments were performed in which annealed crystals were allowed to creep (transiently) for a given period, after which the load was removed (the crystal being otherwise undisturbed) and the crystals allowed to rest at 35°C. After various times, the load was reapplied and the crystals permitted to creep again.

Zinc exhibits recovery at room temperature and one therefore expects the creep rate to be larger after resting than just before, falling eventually to the extrapolation of the original creep rate curve. A series of experiments performed on one crystal, in which the first period of creep was 30 minutes, gave results consistent with these ideas. It was found, however, that the creep rates after rest periods of 1, 2, 10, 50, and 250 minutes were all the same, within experimental error. Thus, Figs. 2 and 3 show the results obtained with rest periods of 1, 10, 50, and 250 minutes. Little difference is apparent. Similar results were obtained with two other crystals which had been allowed to creep for 150 minutes before the rest period (curves 1 and 2 of the figure in reference 16), although here there was a noticeable difference in the effects of a 1-minute rest and a 100-minute rest.

Ten crystals—five grown in nitrogen and five in argon—were each allowed to rest for two days after 150 minutes of creep and then again subjected to the original stress. Two of the nitrogen-grown specimens and two grown in argon appeared to recover completely during a 40-hour rest. This might have been expected on the basis of stress-strain experiments of Haase and Schmid[14] and Cottrell and Aytekin.[15] Two others, argon-grown, did not fully recover their original softness after the rest. The remaining four exhibited the phenomenon termed "rest-hardening" in an earlier report.[16] It was found that although a rest of 1 to 100 minutes produced the limited recovery discussed in the preceeding paragraph, these crystals appeared to have become harder after a rest period of 25–70 hours. The creep rate after resting had an initial value near to that just before the rest period, and the rate very rapidly fell far below the extrapolation of the original log rate—log time curve. This behavior was found to occur consistently in these four crystals. The phenomenon of rest-hardening is thus a characteristic property of a given specimen.

It is important to know whether the first 150 minutes of creep are necessary in order that a crystal show rest-hardening; that is, is prior deformation essential to the production of the effect? With those crystals that displayed the hardening it was found that merely a rest period of several days after annealing did sometimes result in rather low creep rates, although the hardening (decrease in Andrade's β) was generally less than that resulting from a rest following 150 minutes of creep. Moreover, in other tests, a rest of several days produced no hardening at all if a period of creep had not preceded the rest, although these same crystals consistently exhibited the rest-hardening when subjected to the usual procedure of 150 minutes of creep followed by the rest period. One therefore concludes that prior deformation, if not always essential, certainly intensifies the hardening effect of a rest period.

Contraction after Creep

A considerable contraction with time after creep (after effect) is often found with polycrystalline specimens, presumably because of the different amounts of elastic and plastic strain produced in the various crystals of different orientations. With single crystals,

[13] S. Harper and A. H. Cottrell, Proc. Phys. Soc. (London) **B63**, 331 (1950).

[14] O. Haase and E. Schmid, Z. Physik **33**, 416 (1925).

[15] A. H. Cottrell and V. Aytekin, J. Inst. Metals **77**, 389 (1950).

[16] L. Slifkin and W. Kauzmann, Phys. Rev. **78**, 631 (1950).

however, the only observation of an effect related to this of which the authors are aware is that of Sachs and Shoji[17] on brass. In a few instances in the present investigation, contractions of the zinc crystals after creep were measured. The instantaneous contraction after removing the stress generally exceeded the calculated elastic strain (about 10^{-6}) by a factor of 10. No correlation was found between the extent of this excess shringage and the amount of creep that had preceded it. Crystals also sometimes continued to contract over a period of hours, at a rate decreasing with time. The amount of this further contraction varied from zero up to about 10^{-5} cm/cm. These observations were all made at 35°C. No attempt was made to detect or measure any changes in length which might have occurred during the annealing treatment at 210°C.

DISCUSSION

Cottrell and Aytekin[3,15] have recently performed experiments on the creep of zinc single crystals. They found that if the angle α between the c axis and the specimen axis was greater than 45° and if the temperature was less than 120°C, the results agreed with Andrade's equation within experimental error. Otherwise, they sometimes observed accelerating creep which was attributed to a geometric "softening" effect in which lattice rotation increases the resolved shear stress faster than the crystal strain hardens. When Andrade's law was obeyed, β was usually 2 or 3 orders of magnitude larger than K (using the minute as the unit of time). Cottrell and Aytekin's transient creep rates were about 100 times greater than ours.

Our results are somewhat at variance with these. Firstly, we found Andrade's law was only approximately true in that the exponent n of the transient creep equation usually deviated appreciably from the value $\frac{2}{3}$. Cottrell and Aytekin plotted their data as strain *versus* time and picked best values of Andrade's parameters to fit the results, whereas we measured the slope of the log rate—log time plot. The latter procedure probably gives a more critical estimate of n. Secondly, we sometimes observed accelerating creep at a temperature of 35°C and when α was greater than 45°. For α greater than 45°, the lattice rotation due to extension decreases the resolved shear stress. Moreover, the strains involved in our experiments were so small that rotation of lattice planes was negligible. One therefore concludes that the geometric softening hypothesis cannot explain all cases of accelerating creep of single crystals. Thirdly, although in our work, as in that of Cottrell and Aytekin, β/K (in units of $\text{min}^{\frac{1}{3}}$) was very great at lower stresses, we have found that at larger stresses (such that the transient creep parameter a was as large as about $10^{-4}\ \text{min}^{(n-1)}$), β was only an order of magnitude greater than K. (That is, after 20 or 30 minutes of creep both transient and quasi-viscous flows were making comparable contributions to the elongation.) Thus the ratio β/K decreases with increasing β. Since Cottrell and Aytekin worked with higher creep rates than ours and yet found $\beta/K \approx 10^2-10^3$, then either (a) this ratio as a function of strain rate goes through a minimum at some intermediate rate; (b) the mechanisms of plastic flow are different in the two cases because of the different rates involved; (c) their crystals being thin wires while ours were $\frac{1}{4}$-inch rods, perhaps the diameter of the specimen is an important factor; or (d) other effects such as handling, impurities, or surface condition are entering here. The second alternative seems plausible, but since their crystals were wires, the effects of handling may also have been appreciable.

Tyndall[12] has reported that when the angle between the glide direction and the projection of the specimen axis on the glide plane was $\leqslant 2°$, the exponent n in the transient creep equation was equal to 0.5. For our crystal No. 9 (prepared, however, from reagent grade zinc) this angle was 1° and the values of n were not close to 0.5 (see Table I). Tyndall made successive runs by simply increasing the applied stress and without annealing between tests; this would not be expected to give the same results as the method used in this work.

As long as the deformation occurs chiefly by transient creep, many runs can be made on a single crystal with no apparent change in the plastic properties of the specimen by simply annealing after each run. This is significant for two reasons: (a) One can perform a series of experiments in which much of the unpredictable variability among different crystals can be eliminated. The proportionality between γ_0 and β is observed only in such experiments. (b) It appears that the changes which take place in the crystal during transient creep are not permanent and can be removed by annealing at 210°C. If one considers the flow to have its origin in the effects of certain generators, such as of the type proposed by Frank and Read[18] to provide stress-induced production of dislocations in certain slip planes, then these results indicate that the decrease of creep rate with time is caused by a jamming or blocking of the generators and not by their destruction. For if transient creep involved the destruction of successively less potent generators, then one would have to assume that new ones were produced during each anneal. It would then be difficult to account for the good reproducibility of creep tests with the same crystal as compared to the large variation found from specimen to specimen. On the other hand, if the decrease in creep rate with time were due to the accumulation of dislocations or to the jamming effects of vacancies produced by moving dislocations, as Seitz[19] has recently proposed, then the congestion could be removed by annealing and the crystal would be returned to its

[17] G. Sachs and H. Shoji, Z. Physik **45**, 776 (1927).

[18] F. C. Frank and W. T. Read, Jr., Phys. Rev. **79**, 722 (1950).

[19] F. Seitz, Phil. Mag. Supplement **1**, 48 (1952). See also F. Seitz, Phys. Rev. **79**, 1002 (1950); **80**, 239 (1950).

original state. Perhaps the observation that the coarseness of the fluctuations in creep rate appeared to be a characteristic property of a given specimen throughout many tests may also be a result of a distribution of generators which remains essentially unchanged in a series of runs with the crystal.

It is difficult to explain the wide variations in tensile strength found in different specimens which were supposedly prepared in the same manner. J. S. Koehler has pointed out to the authors that the manner in which the specimen is cooled after solidification may have some effect on the distribution of dislocations and other lattice defects. Slight variations in the cooling schedule may thus be an important factor in causing variability among crystals.

It is interesting that surface oxidation should produce such a large increase in tensile strength. This effect suggests that the dislocations whose motions are responsible for the flow must move all of the way to the surface of the crystal. Harper and Cottell,[18] in their study of the effect of surface oxidation on the plasticity of 1-mm zinc single crystal wires found succeedingly smaller effects of the oxide coat at smaller and smaller strain rates, and observed no effect if the strain rate was as low as 10^{-5}/sec. Our observations on $\frac{1}{4}$-inch specimens, however, indicate that an effect exists at still lower strain rates. The conclusions of Harper and Cottrell were based primarily on stress-strain tests. These would not necessarily give the same results as creep tests, which are generally more sensitive to the state of the crystal.

In a series of runs with a given crystal, the initial instantaneous plastic extension is proportional to the transient creep parameter a. It is, of course, possible that this γ_0 is an artifact, due to the shock of loading. This appears improbable, however, for two reasons: (a) The stress was applied by lowering the beam support by means of a manually operated screw. Hence the time of loading, and therefore the shock stress, would be expected to vary considerably from test to test. Since the flow rate increases exponentially with stress, this would produce great variations in γ_0. (b) The shock stress should vary roughly linearly with the load being applied, but the flow rate varies exponentially with stress. If γ_0 were due to the shock load, it would not then be proportional to a.

The existence of this proportionality shows that the instantaneous initial flow and transient creep are related; presumably both are dependent on the distribution of dislocation generators present in the crystal. For example, the initial plastic extension may be the result of unimpeded operation of all generators which are activated by the applied stress. The rapid accumulation of dislocations and vacancies would quickly curb this "instantaneous" flow and transient creep then ensues. Such a model could thus account for the observed proportionality.

The experiments concerned with recovery show that the softening effect per unit time of resting decreases extremely rapidly with time. Cottrell and Aytekin[19] have studied the recovery of zinc crystals by measurement of the decrease in the yield stress of a strain-hardened specimen with time. These stress-strain tests indicated that the degree of recovery increased logarithmically with time, a dependence also found in the study of the recovery of other properties.[20] Such a relation predicts that in our recovery experiments the effects of additional resting should decrease rapidly. Inspection of Figs. 2 and 3 shows that whatever process produced recovery in the first minute of rest has almost ceased to operate after that first minute. One is therefore tempted to infer that the recovery of zinc crystals, deformed as in this work, proceeds in at least two stages: (1) a rapid, limited recovery, completed within one minute. The results of this process are soon exhausted upon further creep. (2) The very slow (at 35°C) process which in several days usually returns the crystal to its original softness. Cherian, Pietrokowsky, and Dorn[21] performed stress-strain experiments in which they found two similar stages in the recovery of polycrystalline aluminum. They termed these *meta-* and *ortho*-recovery, respectively. In Figs. 2 and 3 it is seen that by 20 minutes after the rest period the crystal crept as though there had been no rest period interposed. This suggests that the *meta*-recovery does not involve any large-scale reorganization or healing of the defects which produce strain-hardening. As mentioned above, Seitz has suggested that many vacancies are produced about dislocation generators during plastic flow, that these contribute to strain-hardening, and that their effects should disappear rapidly upon annealing at temperatures at which they are mobile. It may well be, therefore, that this *meta*-recovery is due to clustering and/or diffusion away of such vacancies. The slower, but more effectual, *ortho*-recovery could be the result of the movement of dislocations.

The explanation of the fact that some crystals exhibited rest-hardening after a two-day rest instead of complete resoftening is not clear. This phenomenon is similar to the thermal-hardening, or yield point effect, observed in stress-strain tests of zinc crystals by Orowan[22] and by Wain and Cottrell.[23] In thermal hardening, if, after a stress-strain test, the crystal is allowed to rest for a few days and then again subjected to stress, no flow is observed until the applied stress is increased to a value much greater than the highest stress in the previous test. Once flow starts, the crystal immediately softens. Wain and Cottrell attribute the effect to the migration of dissolved nitrogen atoms to

[20] W. G. Burgers, Proc. Koninkl. Nederland. Akad. Wetenschap. **50**, 452 (1947).
[21] Cherian, Pietrokowsky, and Dorn, Trans. Am. Inst. Met. Engrs. (J. Metals) **185**, 948 (1948).
[22] E. Orowan, Proc Phys. Soc. (London) **52**, 8 (1940).
[23] H. L. Wain and A. H. Cottrell, Proc. Phys. Soc. (London) **B63**, 339 (1950).

dislocations during the rest period, thus anchoring the dislocations until they are later torn loose by a very large stress.

There are two differences between these observations and our work: (a) In our creep tests, reapplication of the original stress after the rest period produced some flow, contrary to Orowan's results, and (b) the crystal did not soften upon further deformation after the rest period. Perhaps the disagreements are due to the fact that the thermal hardening experiments involved stress-strain tests whereas we studied creep. They might also be the result of the much smaller strains and strain rates employed in this work. If the two effects are essentially the same and if the theory of Wain and Cottrell is correct, then the fact that not all crystals exhibited the effect might result from variations in the concentration of dissolved nitrogen in the several specimens. There is some question as to the validity of this argument, however, since one of the argon-grown crystals showed conspicuous rest-hardening, and this crystal was one of the four argon-grown specimens which had low tensile strength, as described above. One cannot, then, attribute both rest-hardening and the higher tensile strength of the other crystals to the presence of traces of nitrogen.

The small total strain, (of the order of 5×10^{-6}) which was produced in our experiments after severe rest-hardening could be the result of a few flow generators which remain active. Another possibility is that it is due to the bowing out of sections of many dislocations which are anchored at various points, a model originally suggested by Koehler[24] to explain some internal friction measurements. Assuming a density of about 5×10^{8} cm of dislocation line per cc, the required strain would be produced if each dislocation moved an average of only 10 atom distances in the slip direction. Thus, a bowing mechanism could account for strains of this magnitude. Such a process produces a homogeneous deformation rather than a flow concentrated in slip bands, but it is doubtful whether one could experimentally assess the homogeneity of the small strain produced after rest-hardening.

[24] J. S. Koehler, Tech. Rep. No. 5, Contract N6ori-47, task order 1 (1950).

Wain and Cottrell's theory of thermal hardening predicts that prior deformation should not be necessary in order that a rest period produce an increase of tensile strength. Indeed, this was verified in stress-strain tests of cadmium crystals by Cottrell and Gibbons.[25] Our results, however, indicate that prior creep intensifies the rest-hardening. Again, if the two effects have the same origin the discrepancy may lie in the different amounts of strain involved. This dependence of hardening on prior creep could be explained by Seitz's theory that vacancies produced during plastic flow will anchor dislocations or will aid in the migration of otherwise immobile foreign atoms to the dislocations by the mechanism suggested by Johnson.[26] Thus, dislocations which would be free after resting without prior deformation could become anchored during a rest after deformation.

The observation that some of the crystals contracted after removing the stress is interesting in view of a recent paper by Kuhlmann[27] on the theory of plastic deformation. According to this theory, during creep dislocations of like sign are piled up behind obstacles by the applied stress. Upon removing the stress, the mutual repulsion of the dislocations causes a redistribution, resulting in a reversal of the flow. Miss Kuhlmann therefore predicts the occurrence of contractions after creep in single crystals, such as we have found. These contractions were of the order of 10^{-5} cm/cm. They could perhaps also be understood in terms of dislocation bows such as are discussed above. After removal of the stress, these bows would tend to retract to form approximately straight lines between the anchors, thereby producing a contraction of the specimen. Presumably, some would have acquired new anchors at the bowed-out sections and could not retract until thermal activation freed them. One would therefore predict still more contraction when the specimens were heated to a higher temperature.

We wish to thank Dr. J. S. Koehler and Dr. F. Seitz for many valuable discussions.

[25] A. H. Cottrell and D. F. Gibbons, Nature **162**, 488 (1948).
[26] R. P. Johnson, Phys. Rev. **56**, 814 (1939).
[27] D. Kuhlmann, Proc. Phys. Soc. (London) **A64**, 140 (1951).

The Theory of an Oscillating Cylinder Viscometer. Part I

ALI A. K. IBRAHIM AND A. M. KABIEL
Physics Department, Farouk Ist University, Alexandria, Egypt

(Received February 27, 1952)

The theoretical distribution of velocity across a liquid (Newtonian) between two coaxial cylinders and the velocity gradient at the inner cylinder is calculated in a new way. The equation used for calculating the coefficient of viscosity of the liquid is also given. The theory agrees with experiment for all values of $\theta_0/\phi_0 \leqslant 1/2$ (where θ_0 is the amplitude of deflection of the inner cylinder and ϕ_0 is the amplitude of the oscillations inexorably imposed on the outer cylinder).

THE THEORY

CONSIDER a liquid between two coaxial cylinders; the inner one is suspended by a torsion wire while the outer is mounted on a table which by means of connecting rods to an electric motor can be made to oscillate through small angular amplitudes.

Since the motion of a liquid is symmetrical about a point and in two dimensions, it follows that the velocity (V) is determined by the equation[1]

$$\frac{1}{\nu}\cdot\frac{dV}{dt}=\frac{d^2V}{dr^2}+\frac{1}{r}\cdot\frac{dV}{dr}-\frac{V}{r^2}, \tag{1}$$

where ν is the kinematic viscosity.

If the sides of the fluid are given an oscillatory torque, therefore,

$$\theta=\phi e^{i\omega t}, \tag{2}$$

where ϕ is a function of r only and ω is the angular velocity of the outer cylinder.

Therefore, we have

$$\frac{d^2\phi}{dr^2}+\frac{3}{r}\cdot\frac{d\phi}{dr}+\frac{2B^2}{i}\cdot\phi=0, \tag{3}$$

where

$$B=(\pi\rho/\eta T)^{\frac{1}{2}},$$

ρ is the density of liquid, η its coefficient of viscosity, and T is the periodic time of the outer cylinder.

Now if

$$b-a=2d,$$

where b is the inner radius of the outer cylinder and a is the radius of the inner cylinder. Therefore

$$r=R+y,$$

where

$$R=(a+b)/2$$

and

$$y \text{ varies between } \pm d.$$

Therefore, for a first approximation which is quite convenient, Eq. (3) becomes

$$\frac{d^2\phi}{dy^2}+\frac{3}{R}\cdot\frac{d\phi}{dy}+\frac{2B^2}{i}\cdot\phi=0. \tag{4}$$

[1] Lamb, *Hydrodynamics* (Dover Publications, New York, 1945).

The general solution of Eq. (4) is[2]

$$\phi=C_1\exp\{(-3/2R+C+iD)y\}+C_2\exp\{(-3/2R-C-iD)y\}, \tag{5}$$

where

$$\left.\begin{aligned}C&=\{[9/8R^2+(81/64R^4+B^4)^{\frac{1}{2}}]\}^{\frac{1}{2}}\\ D&=\{[-9/8R^2+(81/64R^4+B^4)^{\frac{1}{2}}]\}^{\frac{1}{2}}\end{aligned}\right\} \tag{6}$$

and

C_1 and C_2 are arbitrary constants.

Thus

$$\theta=C_1\exp\{(-3/2R+C+iD)y+i\omega t\}+C_2\exp\{(-3/2R-C-iD)y+i\omega t\}$$

$$\dot{\theta}=i\omega[C_1\exp\{(-3/2R+C+iD)y+i\omega t\}+C_2\exp\{(-3/2R-C-iD)y+i\omega t\}],$$

and

$$d\dot{\theta}/dy=i\omega[C_1(-3/2R+C+iD)\exp\{(-3/2R+C+iD)y+i\omega t\}+C_2(-3/2R-C-iD)\times\exp\{(-3/2R-C-iD)y+i\omega t\}]. \tag{7}$$

The Equation of Motion of the Inner Cylinder

Consider the forces on the inner cylinder.

(A) The viscous forces over the curved surface of the inner cylinder.

If L is the length of the inner cylinder from the surface of the liquid to the lower extremity, therefore,

$$F=\eta\left[\left(\frac{dv}{dy}\right)_{y=-d}-\frac{(V)_{-d}}{a}\right]\cdot 2\pi aL,$$

where F[3] is the force due to the viscosity over the curved surface of the inner cylinder.

The moment of this force about the common axis is C'', where

$$C''=\eta\left[\left(\frac{dv}{dy}\right)_{y=-d}-\frac{(V)_{-d}}{a}\right]2\pi a^2L$$

since

$$V=r\dot{\theta}.$$

[2] Lamb, *Differential Calculus* (Cambridge University Press, New York, 1949).

[3] Basset, A. B., *Hydrodynamics*, 1900.

Therefore,

$$C''=2\pi La^2\eta\left(\frac{d\dot{\theta}}{dy}\right)_{y=-d}. \tag{8}$$

(B) The restoring force in the supporting wire of the inner cylinder.

If τ is restoring couple per unit angular twist, therefore, the restoring couple for angular twist θ_0 is C', where

$$C'=\tau\theta_0. \tag{9}$$

Since the couples C'' and C' are opposite in direction, therefore the equation of motion of the inner cylinder becomes

$$I\ddot{\theta}=-\tau\theta_0+2\pi L\eta a^2\left(\frac{d\dot{\theta}}{dy}\right)_{y=-d}. \tag{10}$$

Thus,

$$\theta_0=\frac{2\pi L\eta a^3}{\tau-I\omega^2}\left(\frac{d\dot{\theta}}{dy}\right)_{y=-d}. \tag{11}$$

and

$$\begin{aligned}(\dot{\theta})_{-d}=\frac{2\pi L\eta a^3\omega^2}{I\omega^2-\tau}[&C_1(-3/2R+C+iD)\\&\times\exp\{(-3/2R+C+iD)(-d)+i\omega t\}\\&+C_2(-3/2R-C-iD)\\&\times\exp\{(-3/2R-C-iD)(-d)+i\omega t\}],\end{aligned} \tag{12}$$

where $(\dot{\theta})_{-d}$ is the angular velocity of the liquid in the neighborhood of the inner cylinder.

The Boundary Conditions

The next step in solving the problem is to determine the constants C_1 and C_2. To do so we make use of the following conditions which may be associated with the practical side of the problem.

When

$$y=+d,$$
$$\dot{\theta}=i\omega\phi_0e^{i\omega t},$$

where ϕ_0 is the amplitude of the oscillations inexorably imposed on the outer cylinder, and when

$$y=-d,$$
$$\dot{\theta}=(\dot{\theta})_{-d}.$$

Therefore,

$$\left.\begin{aligned}C_1&=\frac{E'\phi_0e^{i\omega t}}{FE'-F'E}\\ \text{and}\quad C_2&=\frac{-F'\phi_0e^{i\omega t}}{FE'-F'E}\end{aligned}\right\}, \tag{13}$$

where

$$\begin{aligned}F&=\exp\{(-3/2R+C+iD)d+i\omega t\}\\E&=\exp\{(-3/2R-C-iD)d+i\omega t\}\\F'&=\exp\{(-3/2R+C+iD)(-d)+i\omega t\}\\&\quad\times[\gamma(-3/2R+C+iD)-i\omega]\\E'&=\exp\{(-3/2R-C-iD)(-d)+i\omega t\}\\&\quad\times[\gamma(-3/2R-C-iD)-i\omega]\end{aligned}$$

and

$$\gamma=2\pi L\eta a^3\omega^2/(I\omega^2-\tau).$$

Therefore,

$$\begin{aligned}\dot{\theta}=\frac{i\omega\phi_0e^{i\omega t}}{FE'-F'E}[&E'\exp\{(-3/2R+C+iD)y+i\omega t\}\\&-F'\exp\{(-3/2R-C-iD)y+i\omega t\}]\end{aligned} \tag{14}$$

and

$$\begin{aligned}\left(\frac{d\dot{\theta}}{dy}\right)_{y=-d}=\frac{i\omega\phi_0e^{i\omega t}}{FE'-F'E}[&E'(-3/2R+C+iD)\\&\times\exp\{(-3/2R+C+iD)(-d)+i\omega t\}\\&-F'(-3/2R-C-iD)\\&\times\exp\{(-3/2R-C-iD)(-d)+i\omega t\}].\end{aligned} \tag{15}$$

In solving Eq. (14) we neglect the higher powers of d^2 and γ^2 since they are very small compared with either τ in the case of low frequency or I in the case of high frequency. Therefore we have

$$\dot{\theta}_{\text{amplitude}}=\frac{2\omega^3d(d+y)\{1-[(3/R)-(\gamma\rho/\eta)][(d+y/2)]\}\{1+d[(\gamma\rho/\eta)-(3/R)]\}+\omega\gamma^2}{\gamma^2+4\omega^2d^2\left[1+d\left(\frac{2\pi\rho La^3\omega^2}{I\omega^2-\tau}-\frac{3}{R}\right)\right]^2}\phi_0, \tag{16}$$

also if $\gamma^2\to0$

$$\dot{\theta}_A=\frac{\omega(d+y)\{1+[(\gamma\rho/\eta)-(3/R)](d+y/2)\}}{2d\{1+[(\gamma\rho/\eta)-(3/R)]d\}}\phi_0$$

$$(\dot{\theta})_{-d}=\lim_{\gamma^2\to0}\frac{\omega\gamma^2\phi_0}{4\omega^2d^2\left[1+d\left(\frac{2\pi\rho La^3\omega^2}{I\omega^2-\tau}-\frac{3}{R}\right)\right]^2+\gamma^2}=0 \tag{17}$$

and

$$\left(\frac{d\dot{\theta}}{dy}\right)_{y=-d}=\frac{\omega\phi_0}{2d\left[1+d\left(\frac{2\pi\rho La^3\omega^2}{I\omega^2-\tau}-\frac{3}{R}\right)\right]}. \tag{18}$$

Therefore, from (18) and (11) we have

$$\eta=\frac{b-a}{L}\cdot\frac{\theta_0}{2\pi}\cdot\frac{1+\frac{b-a}{2}\left(\frac{2\pi L\rho a^3\omega^2}{I\omega^2-\tau}-\frac{3}{R}\right)}{a^3\phi_0}\cdot\frac{I\omega^2-\tau}{\omega}. \tag{19}$$

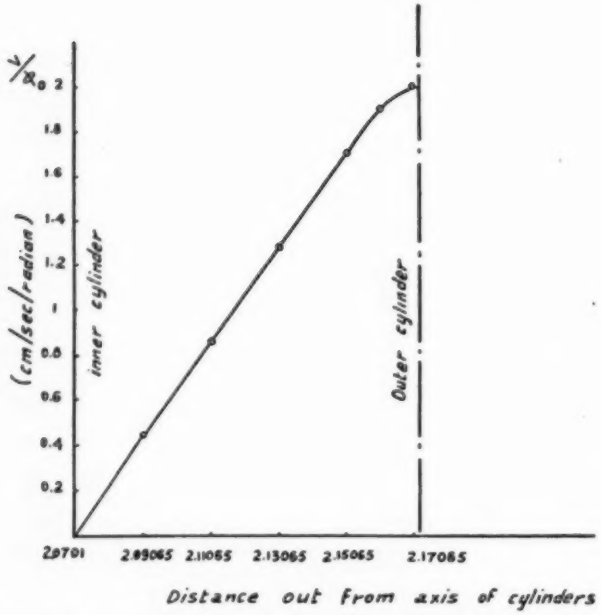


FIG. 1. Velocity gradient in liquid paraffin ($\rho=0.877$ gm cm^{-3} at 25°C) at frequency ($\omega=0.982$ sec^{-1}) of outer cylinder.

If τ is very small compared with $I\omega^2$, we have

$$\eta=\frac{\theta_0}{2\pi\phi_0}\cdot\frac{b-a}{L}\cdot\frac{I\omega\left[1+\frac{b-a}{2}\left(\frac{2\pi\rho La^3}{I}-\frac{3}{R}\right)\right]}{a^3}. \quad (20)$$

Therefore Eq. (19) or the approximate one, Eq. (20), is the required equation for calculating the coefficient of viscosity of the liquid.

If θ_0 is determined by a mirror attached to inner cylinder and a scale provided with a telescope for observation at a distance M from the mirror,

$$\theta_0=\frac{S}{2M},$$

where S is the observed deflection of the scale image. Therefore

$$\eta=\frac{1}{4\pi}\cdot\frac{b-a}{\phi_0 L}\cdot\frac{S}{M}\cdot\frac{I\omega\left[1+\frac{b-a}{2}\left(\frac{2\pi L\rho a^3}{I}-\frac{3}{R}\right)\right]}{a^3}. \quad (21)$$

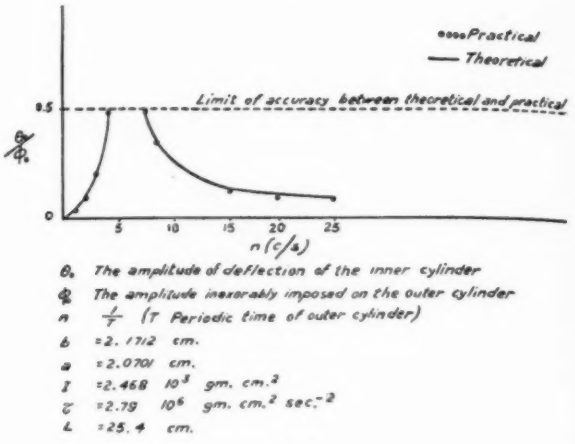


FIG. 2.

Figure 1 shows the theoretical variation of velocity of the liquid between cylinders with distance between the two cylinders.

Experimental Verification

From Fig. 2 we see that for a Newtonian liquid (liquid paraffin B.P. of density 0.877 gm/ml and viscosity 1.76 poises at 25°C) the experimental results[4] of Oldroyd, Strawbridge, and Toms agree accurately with our theoretical equation,

$$\eta=\frac{b-a}{L}\cdot\frac{\theta_0}{2\pi}\cdot\frac{1+\frac{b-a}{2}\left(\frac{2\pi L\rho a^3\omega^2}{I\omega^2-\tau}-\frac{3}{R}\right)}{a^3\phi_0}\cdot\frac{I\omega^2-\tau}{\omega} \quad (19)$$

for all values of $\theta_0/\theta_0\leq\frac{1}{2}$. The agreement appears to justify the neglect of end effects in our theory, as well as indicating the limitation in our proposed theory.

ACKNOWLEDGMENT

The authors are greatly indebted and wish to express gratitude to Professor M. A. Elsherbini for his great interest and advice.

Also the discussions with Dr. Meksyn are greatly acknowledged.

[4] Oldroyd, Strawbridge, and Toms, Proc. Phys. Soc. (London) **B64**, 44 (1951).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Measurements of the Physical Properties of Active Nitrogen

JAMES M. BENSON*
Hastings Instrument Company, Inc., Hampton, Virginia
(Received March 3, 1952)

Measurements were made of the heat released during decay of the afterglow, the viscosity, and the free electron density in active nitrogen. Calorimeter measurements of the heat released during decay of the afterglow amounted to as much as 22.3 calories per gram of the gas, corresponding to an excitation of one part in 350 of the molecules present to an energy of 9.6 electron volts. The viscosity of the active nitrogen was found to be the same as that of the unactivated nitrogen within one-half of one percent. Ionization in a stream of active nitrogen is shown to be incidental rather than essential to the production of the Lewis-Rayleigh afterglow. Microwave measurements of the free electron density in a stream of active nitrogen indicated one free electron for each 2.3×10^8 molecules in the stream. Contrary to Mitra's theory, probes in the stream were found to respond to carriers of negative sign rather than positive ions. Space discharges in an expanding stream of active nitrogen are attributed to free electrons moving with the stream. In the use of active nitrogen for studies of low density flow, by observation of the afterglow, disturbing effects will occur because of heating and ionization, but the disturbances appear not to be so large as to interfere with its use.

INTRODUCTION

ACTIVE nitrogen is a form of nitrogen that results from suitable excitation by a strong electric discharge at low pressures. The visible afterglow that persists after the discharge is stopped was studied by E. P. Lewis as early as 1900. The second Lord Rayleigh (R. J. Strutt) observed that the excited gas is chemically active.[1] He also observed that when any one of certain other elements, for example, sodium or mercury, is introduced into a stream of active nitrogen, the yellow afterglow is replaced by emission characteristic of the element added to the stream. A striking feature of the spectra of elements added to active nitrogen is that they are limited to selected members of complete spectra and are representative of low temperature emission.

The spectrum of the afterglow of active nitrogen is markedly different from that of the glow discharge. The glow discharge of nitrogen includes many band systems, while the spectrum of the pure afterglow of active nitrogen is restricted to the first positive system. Furthermore, certain members of the first positive system are selectively enhanced in the afterglow to produce an intensity distribution different from that of the first positive system in the nitrogen glow discharge.

Mitra[2] has summarized the large amount of experimental data, including extensive results obtained by Rayleigh,[3] on the chemical and physical properties of active nitrogen up to 1945. It is clearly established that the yellow afterglow characteristic of active nitrogen contains no atomic lines and contains no band system which cannot be ascribed to the normal nitrogen molecule. Kaplan[4] has named the yellow afterglow of active nitrogen the "Lewis-Rayleigh" afterglow to distinguish it from other afterglows in nitrogen. The emission of the afterglow takes place by allowed transitions between electronic states of the molecule; therefore, the long lifetime of the afterglow is not to be attributed to the persistence of excited states of the upper levels from which radiating transitions occur.

Theories that assume the active centers to be metastable molecules readily account for the energy as determined spectroscopically. However, Mitra has pointed out that metastable molecules do not account for the energy of 12.9 electron volts that were measured by Rayleigh in experiments on the heating of metal foil placed in active nitrogen. Mitra has proposed another theory in which he assumes the active centers to be singly ionized nitrogen molecules having the ionization energy of 15.58 electron volts. Other theories have been proposed in which it is assumed that both atomic nitrogen and metastable molecules are present in active nitrogen.

The present evaluation of the physical properties of streaming active nitrogen provides new information for examining the various theories of the afterglow. The results of the tests also clarify some of the more important questions that arise in applications of the afterglow to studies in compressible flow at low densities.

POPULATION AND AVERAGE ENERGY OF ACTIVE CENTERS

Estimates from Rayleigh's Measurements

Early measurements by Rayleigh[1] of the absorption of active nitrogen by phosphorous indicated that about one part in 210 of streaming nitrogen became activated by an electric discharge. Subsequent measurements

* This report includes a portion of the material submitted to the University of Tennessee in the form of a Ph.D. thesis. The experimental work was carried out while the writer was employed at the Langley Aeronautical Laboratory of the NACA, and reported in NACA TN 2293, February 1951.

[1] R. J. Strutt, Proc. Roy. Soc. (London), **A85**, 219 (1911); **A86**, 56 (1911).

[2] S. K. Mitra, *Active Nitrogen—A New Theory* (Indian Press, Ltd., Calcutta, 1945).

[3] Lord Rayleigh, Proc. Roy. Soc. (London), **A151**, 567 (1935), **176**, 1–16 (1940); **180**, 123, 140 (1942).

[4] Joseph Kaplan, Phys. Rev. **54**, 176 (1938).

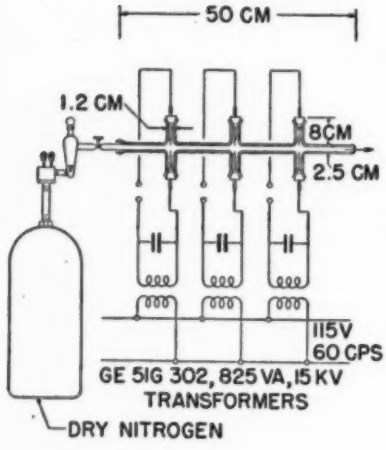


FIG. 1. Diagram of apparatus for exciting the stream of active nitrogen.

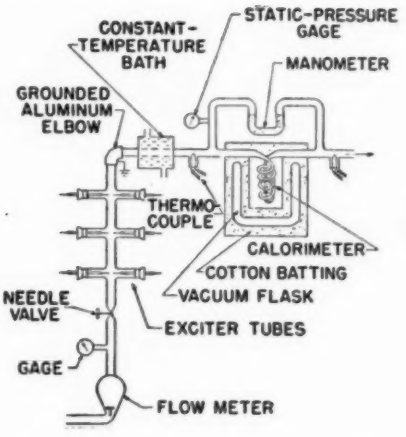


FIG. 2. Arrangement of viscosimeter and calorimeter.

by him of the formation of N_2O_3 from nitric oxide and active nitrogen indicated that one part in 40 was activated.

In measurements of the visible flux from a known mass of nitrogen when excited to produce as much afterglow as could readily be obtained, Rayleigh found that the total flux in the visible region was roughly equivalent to the emission of one quantum for each 1300 molecules present. Additional emission is known to occur in the infrared and ultraviolet regions so that it is to be expected that more than 1 molecule in 1300 is activated.

Rayleigh's measurements of the heating of gold foils indicated that an extraordinary amount of energy could be recovered from the active nitrogen. He recovered in the form of heat on a piece of gold sheet as much as 12.9 electron volts of energy for each molecule that passed through the exciter. As Rayleigh pointed out, this amount of energy exceeds that available from all the gas that passes through the apparatus due to dissociation in the exciter and recombination on the gold sheet. If the gas were almost completely ionized as postulated by Mitra, there would be enough energy available from recombination on the metal surface. More recent experiments by Rayleigh[5] resulted in still larger amounts of energy recovered on metal wires.

Calorimeter Measurements of Heat Released by Active Nitrogen

A stream of nitrogen was excited as illustrated in Fig. 1. The active nitrogen was passed through a calorimeter, as shown in Fig. 2, to measure the heat released from the stream. The intensity of the glow and the pressure were measured at the intake and the exhaust ends of the calorimeter. The afterglow was found to decay to approximately 1/20 in passing through the calorimeter during a representative test run with an aluminum coil in the calorimeter; this decay indicated that the excitation responsible for the afterglow also decayed to a small fraction of its initial value. It is to be expected then that the heat recovered by the calorimeter during the decay of the afterglow would be indicative of the energy carried in latent form by the active nitrogen. This expectation is supported by spectroscopic studies reported by Kaplan[6] indicating that the spectrum of the Lewis-Rayleigh afterglow does not appear to change in distribution of intensity as it decays.

Rayleigh's observation that copper was especially effective in quenching the afterglow suggested that a coil of copper tubing in the calorimeter might recover more energy from the active nitrogen than one of aluminum. The glow was faintly visible in the exhaust from an aluminum coil, but no glow could be detected in the exhaust from the copper. The results of the run with a copper coil when both the aluminum elbow and the copper coil were grounded are given in Table I.

As a check on the effect of the fluid friction and expansion of the gas in the coil, a run was made with the flow adjusted as usual, but with no electrical excitation of the nitrogen. In one hour, the temperature of the calorimeter changed less than 0.05°C, indicating an effect of less than 5 percent of the rise due to the active nitrogen.

The results showed that the heat recovery is measurably increased by ion currents to the coil and by more effective quenching of the glow when copper is used instead of aluminum.

The largest amount of energy recovery, 50.6 calories per minute with the grounded copper coil, may be used in an estimate of the maximum average energy release to be expected under the conditions of the ex-

TABLE I. Measurements of heat released by active nitrogen.

Pressure at intake, millimeters of mercury	9.7
Pressure at exhaust, millimeters of mercury	2.9
Mass flow, grams per minute	2.27
Heat flow to calorimeter, cal/min	52.3
Heat flow indicated by temperature drop of stream, cal/min	1.7
Net flow ascribed to decay of active nitrogen, cal/min	50.6

[5] Lord Rayleigh, Proc. Roy. Soc. (London), **A189**, 296 (1947).

[6] Joseph Kaplan, Phys. Rev. **73**, 494 (1948).

periment. This is equivalent to 22.3 calories per gram which, if released adiabatically in a stream tube, would raise the static temperature of the stream by 90°C. This value is also equivalent to 0.027 electron volt per molecule of nitrogen. This is only about 1/400 the maximum value measured by Rayleigh in his first series of measurements of heating effects.

In an effort to examine the heating effects observed by Rayleigh, a piece of gold foil 1.2 centimeters square by 0.1 millimeter was placed in the stream of active nitrogen. It was located about 20 centimeters from the nearest of the exciting discharges. No visible incandescence due to active nitrogen was obtained at any pressure or flow rate available with the apparatus on hand. Then the foil was carefully cleaned and raised to incandescence by using it as one electrode of a high voltage discharge. As soon as the discharge to the gold was stopped, the incandescence faded out even though the exciter discharge remained in operation and the afterglow continued bright. No further effort was made to duplicate the heating effect observed by Rayleigh, since it appeared that the effect was not an essential characteristic of a stream having a bright afterglow.

The calorimeter measurements in these tests show that the population of active centers is very much less than the indications obtained by Rayleigh in his experiments on the heating of gold sheets. The heating effects observed by Rayleigh may have been largely due to heating of the gold sheet by cathode rays as pointed out by Gaydon. The calorimeter measurements, then, are considered to be a more reliable indication of the population of active centers.

From the observed heat recovery of 22.3 calories per gram with a grounded copper coil, an average value of 0.027 electron volt is indicated per molecule in the stream. If the value of 9.6 electron volts is taken as the required excitation to produce afterglow, the indication is that the fraction of molecules excited to this level is of the order of one part in 350. This ratio may be compared with Rayleigh's estimate of one part in 40 from chemical activity and one part in 1300 from the intensity of the visible afterglow. Differences in the amount of excitation are to be expected in the different experiments. In the present investigation, the population of active centers that was obtained is not considered to be the maximum possible. Rayleigh considered that the afterglow obtained by him with much lower rates of flow approached the maximum that could reasonably be expected.

VISCOSITY OF ACTIVE NITROGEN

Mitra's theory, in which he postulates the presence of long-lived positive molecular ions, was presented in part to account for the value of 12.9 electron volts obtained by Rayleigh in heating effects on gold sheet. If this view is correct, a stream of active nitrogen would have an extraordinarily high concentration of ions. In the extreme case cited by Rayleigh, more than 80 percent of the molecules would be ionized. If this amount of ionization is present, it is reasonable to question whether the viscosity of the active gas was very different from that of normal nitrogen. This question is particularly pertinent if the active nitrogen is to be used in visualizing flow. The possibility that the viscosity of active nitrogen differed from that of normal nitrogen was investigated experimentally by measurement of the pressure drop across the calorimeter. Nitrogen was pumped through the exciter and the calorimeter at a constant rate to observe whether any change in viscosity occurred when the exciter was turned on or off. Photoelectric measurements of the brightness of the afterglow indicated that the intensity per unit mass of active nitrogen was at least one-half as great as obtained by Rayleigh. The measurements showed no change in viscosity with excitation on and off. The apparatus was sufficiently sensitive to detect a change as small as one-half of 1 percent of the absolute viscosity.

ELECTRICAL PROPERTIES OF ACTIVE NITROGEN

Ionization

Probe measurements of ionization by Rayleigh and others are conveniently summarized by Mitra as follows:

"1. Ionization is always associated with the glow; the decrease of one is accompanied by decrease of the other. The rates are, however, not the same. The glow intensity decreases more rapidly than the ionization.

"2. The glow and the ionization persist even after the glowing gas from the discharge tube has been drawn through the space between a pair of electrodes across which a saturation current is flowing. (This is done with a view to "trap" the ions which may otherwise be expected to find their way from the discharge tube to the afterglow tube.)

"3. The ionization current between a pair of electrodes is nearly independent of the nature of metal surface. There is exception in the case of copper oxide for which the ionization current is perceptibly higher. (Copper oxide is also highly destructive of the glow.)

"4. The saturation current flowing between a pair of electrodes immersed in the glowing gas is proportional to the area of the negative electrode.

"5. The ionization is not due to emission of electrons from the surface of the metallic electrodes or to photoionization in the volume of the gas.

"6. Introduction of neutral N_2 into the glowing gas increases the ionization current."

Mitra discusses ionization measurements in relation to his postulate that the walls of a vessel containing active nitrogen become conditioned in some way "so that only a very small fraction of the impinging electrons attach themselves to the surface to form surface charge." From this viewpoint is developed the con-

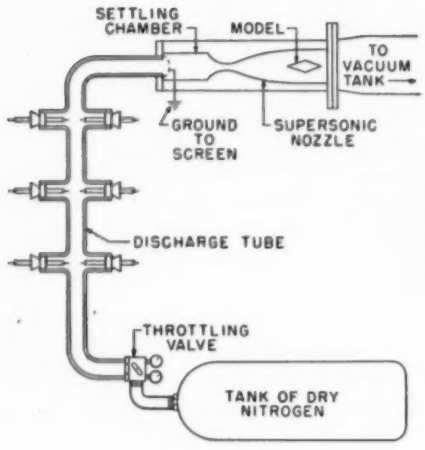


FIG. 3. Diagram of apparatus for producing afterglow in a supersonic stream of nitrogen.

clusion that probes in the active nitrogen will absorb only a small fraction of electrons and ions that strike them and, hence, that probe measurements lead to too small an estimate of the ionization present in active nitrogen.

The results of an experiment by Jackson and Broadway[7] appear to be in disagreement with Mitra's theory that the active centers are charged particles. They applied the Stern-Gerlach method to a study of active nitrogen. A normal Stern-Gerlach pattern was observed with no evidence being noted that the active particles carried an electrical charge. From the description of their apparatus and method, it appears that if the active nitrogen consisted of charged particles as proposed by Mitra, the pattern would have been conspicuously deflected and distorted.

Corona Discharges in the Stream

In a preliminary investigation of afterglow,[8] corona discharges were observed in certain regions of supersonic flow of active nitrogen. The apparatus in which the observations were made is illustrated in Fig. 3. Figure 4 is a photograph of the flow of active nitrogen over a double wedge of aluminum at a Mach number of 2.6. The corona discharge is to be seen in the region of the trailing edge of the model. The corona disappeared if the flow upstream of the nozzle passed through a grounded length of electrically conducting pipe or through a grounded wire mesh.

The space discharge that occurs in the nozzle without a metallic model is shown in Fig. 5. Figure 5(a) is a photograph of the afterglow with no corona in the stream. The flow is directed from left to right. The patterns downstream of the minimum section reveal the dimmer regions of low pressure separated by brighter regions of high pressure resulting from waves being reflected from the boundaries of the stream. Figure 5(b)

[7] L. C. Jackson and L. F. Broadway, Proc. Roy. Soc. (London) A127, 678 (1930).

[8] Thomas W. Williams and James M. Benson, "Preliminary Investigation of the use of afterglow for visualizing low-density compressible flows," NACA TN 1900 (1949).

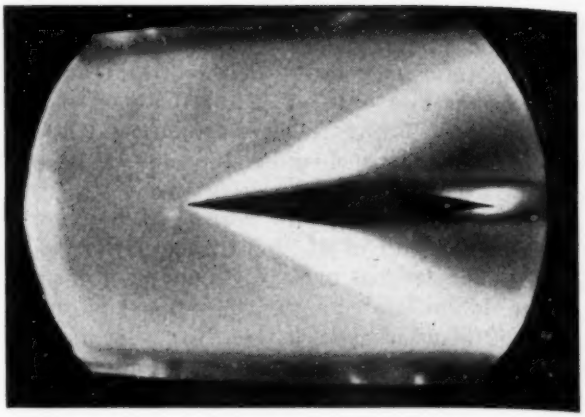
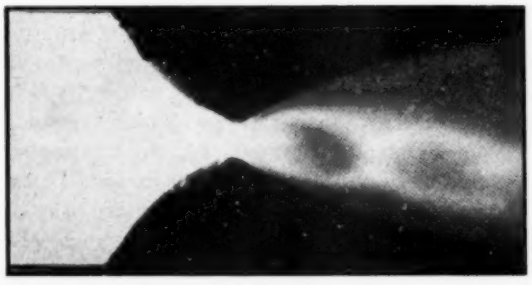
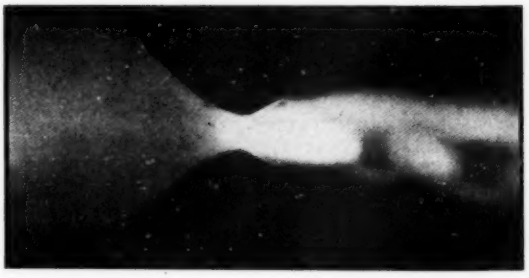


FIG. 4. Photograph of flow over a double wedge showing a corona discharge in the low pressure region around the trailing edge. Mach. number, 2.6.

is a photograph of the flow under the same conditions, except that the ground was disconnected and the space discharge occurs at several places in the nozzle. Figure 5(a) was exposed to show the afterglow patterns in the expanding section of the nozzle. It will be observed that the brighter area in the upstream end of the nozzle is over-exposed. Figure 5(b) was exposed a shorter time to show the corona patterns which are much brighter than the afterglow. This difference in exposure is evident from a comparison of the upstream end of the nozzle in the two photographs. Photoelectric measurements of the intensity of the afterglow in the regions between the screen and the space discharge showed no change when the screen was grounded to remove the



(a)



(b)

FIG. 5. Photograph of separated flow through a supersonic nozzle. (a) With ion trap. (b) Without ion trap.

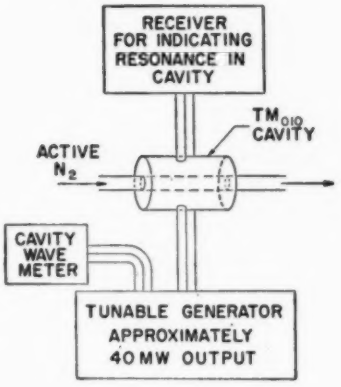


FIG. 6. Arrangement of apparatus for microwave measurements of electron density.

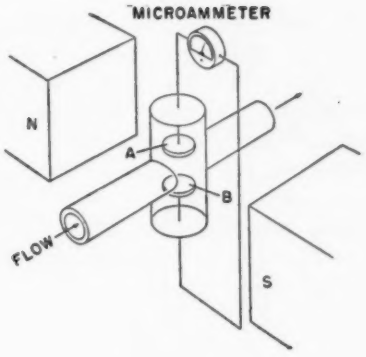


FIG. 7. Arrangement of electrodes in stream near the resonant cavity. With the magnetic field and stream directed as shown, corona formed on electrode A and electrons flowed through the meter from A to B.

discharge. With the space discharge localized in the nozzle, the intensity of the afterglow in the exhaust line was less with the discharge taking place than when the discharge was prevented by grounding the screen. This is in agreement with the observation by Kaplan that a weak, electrodeless discharge in active nitrogen, such as that caused by a Tesla coil, quenches the afterglow.

Consideration has been given to the possibility that the space discharge illustrated in Fig. 5(b) may have resulted from an electrical discharge between the exciter system and the ground. When a magnetic field was applied across the stream at the corona, the glow discharge, but not the afterglow patterns in the stream, was conspicuously deflected in a manner to be expected of negative charges moving with the stream. Since the exciter was powered by alternating current and the deflection was unidirectional, the space discharge in the stream appears not to be the result of a discharge current from exciter to ground.

The observation that removal of charged particles from the stream had no effect on the afterglow was further confirmed by comparing the spectrum of the glow with and without the screen grounded. The spectrum of the afterglow obtained with a small quartz spectrograph (Gaertner model L234–150) on Eastman 103-G plates showed no difference in distribution of intensity or exposure with the screen grounded and ungrounded. Thus, it appears that the density of electric charges in the stream may vary widely without affecting the afterglow except that, if the space charge becomes sufficient to cause a discharge, the afterglow is quenched.

Microwave Measurements

In view of the questions that have been raised, especially by Mitra's theory, regarding the data on ion density obtained with probes, measurements of electron density were carried out by the use of a resonant cavity, as illustrated in Fig. 6. Active nitrogen flowed through the "VYCOR" glass tube concentric with the cavity. The method for measuring the free electron density was the same as described by Rose and coworkers.[9] The electron density measured in this way depended upon the distance from the exciter to the resonant cavity and whether or not there was a grounded length of pipe ahead of the cavity. With the cavity only about 12 inches from the nearest of the three exciting discharges, the effects were too erratic to allow a reading of the resonant wavelength. When an intervening length of 12 inches of aluminum tubing between exciter and cavity was grounded, the resonant peak of the cavity was easily read, showing a resonant frequency of 2601.87 megacycles with the exciter on and the nitrogen at a pressure of 8.0 millimeters. With the exciter off, the resonant frequency was 2601.15 megacycles. The shift of 0.72 megacycle represented an electron density of 7.7×10^8 free electrons per cubic centimeter. It is estimated from thermocouple measurements of temperature of the stream in a similar arrangement that the temperature was roughly 400°K, from which it is estimated that the stream density was 1.9×10^{17} molecules per cubic centimeter. The shift in resonant frequency thus indicates an electron density of about one free electron for each 2.5×10^8 molecules in the stream. The results for additional runs showed that the free electron density increased as the pressure increased up to 12 millimeters of mercury and decreased at higher pressures.

The measured values of the free electron density in active nitrogen are notably high, and it is not surprising that space discharges are observed in a stream having this large a concentration of free electrons. Even so, the concentration is only about 1 free electron for each 2.3×10^8 molecules of nitrogen in the stream, which is much less than the population of active centers indicated by the calorimeter measurements.

Effects of a Magnetic Field on Measurements in the Stream

Brief observations were made of the ion currents obtainable with aluminum electrodes inserted in the stream of active nitrogen. The electrodes were flat cylindrical disks, 1/4 inch in diameter and 1/8 inch thick with embedded thermocouples of iron constantan.

[9] Rose, Kerr, Biondi, Everhart, and Brown, "Methods of Measuring the Properties of Ionized Gases at Microwave Frequencies," M.I.T. Tech. Rep. No. 140. Oct. 17, 1949.

The ion currents were measured by means of a direct-current microammeter in the arrangement illustrated in Fig. 7. With the electrodes, *A* and *B*, approximately flush with the walls of the glass tubing through which the active nitrogen flowed, no current was measured when the circuit was insulated from the ground. It was found that a current of several microamperes could be obtained when a magnetic field of a few hundred gauss was applied transverse to both the stream and a line joining the electrodes. A corona discharge was also observed in the vicinity of one electrode but not the other. The corona shifted from one to the other, and the direction of the current reversed when the magnetic field was reversed. The corona appeared always at the electrode toward which streaming electrons would be deflected by the magnetic field. The thermocouples showed this electrode to be of the order of 10°C warmer than the opposite one.

Effects similar to that caused by the magnetic field were observed when one electrode was moved to the center of the stream and the other withdrawn about $\frac{1}{4}$ inch into the side arm. The current obtained in this way was directed as if negative charges were being collected from the stream. The current obtainable in this way with the magnetic field diminished but did not become zero when a section of the pipe through which the active nitrogen flowed was grounded upstream of the electrodes. Microwave measurements at the same time with the resonant cavity showed that electron density in the stream a few inches ahead of the pair of electrodes also decreased when the pipe was grounded. The decrease indicated by the microwave measurement was roughly proportional to the decrease in the ion current measured between the electrodes.

These results are to be expected, and they indicate that the probes may be used in active nitrogen in a qualitative way. In contrast, Mitra's theory assumes that active nitrogen conditions the electrode surfaces in such a way as to reflect electrons and allow free access of the positive ions to the electrodes. From this, Mitra inferred that electrodes in active nitrogen measured more current due to positive ions than to free electrons. His view disagrees with the present results, which indicate that the electrodes respond to negative rather than positive charges in the stream.

DISCUSSION OF RESULTS

Population of Active Centers

The present measurements of heat released during decay of the afterglow indicated a population of active centers about $\frac{1}{6}$ the maximum that Rayleigh measured by chemical methods. The calorimeter measurements indicated that the population of active centers was approximately one in 240. This is of the same order as Rayleigh's chemical measurements of the population but is very much less than the values he obtained from the heating of metal sheets that he exposed to the active nitrogen.

Ionization

The space discharges in the stream, apparently caused by free electrons moving with the active nitrogen, could be deflected strongly by magnetic fields. In contrast, the glowing stream of active nitrogen showed no response to magnetic fields transverse to the stream. The Stern-Gerlach experiment performed by Jackson and Broadway to measure the magnetic moment of active nitrogen also reported no deflection of the jet. If the jet of active nitrogen in their experiment had been composed of molecular ions as proposed by Mitra, a deflection of the jet should have been observed in addition to the splitting due to the molecular magnetic moment. The evidence then favors the view that ionization effects in the active nitrogen are the results of free electrons rather than positive ions in the stream.

The free electron density measured in the stream appears to be high enough to account for the glow discharges in the stream at regions of low pressure. In support of this view, it may be mentioned that active nitrogen is considered to be excited to states well above the ground state, and the electric field required to cause a glow discharge is less in active nitrogen than normal nitrogen for that reason. It is presumed, then, that the active nitrogen breaks down into a glow discharge in those regions of expanding flow where the electric field, owing to the streaming electrons, becomes sufficiently large.

Observations in the present tests agreed with those of previous tests by Rayleigh and others that the removal of ions from active nitrogen had no effect on the intensity or spectral distribution of the afterglow. The questions raised by Mitra regarding the use of probes in the earlier work to measure ionization were circumvented in the present tests by the use of a resonant cavity to determine that the ion trap was effective in reducing substantially the free electron density in the stream. The present results support the views of Gaydon, Constantinides, and others that ionization in a stream of active nitrogen is incidental rather than essential to the production of afterglow.

Applications of Nitrogen Afterglow to Low Density Flow

The preliminary investigation[8] of nitrogen afterglow in visualizing low density supersonic flow, demonstrated that sharply defined changes in intensity coincided with shock waves. Questions were raised as to the extent to which the aerodynamic properties of active nitrogen would depart from those of normal nitrogen and air.

The results of the viscosimeter and calorimeter tests show that the viscosity and temperature of streaming active nitrogen are not substantially disturbed by

processes associated with the afterglow. The viscosity was found not to change within the experimental error of ½ of 1 percent. The temperature was measurably affected but by amounts that are believed to be acceptable in experimental investigations of low density flow.

The unusually high density of free electrons in a stream of active nitrogen should be given some further consideration in applications of the afterglow to low density flow. Although the viscous forces did not appear to be affected, it is possible that disturbing effects may be found under conditions differing from those in the present tests.

The calorimeter measurements indicated that a temperature rise of as much as 90°C may be expected to occur in a stream while the active nitrogen is returning to its normal state. It is reasonable to question whether additional energy remaining after the decay of the visible afterglow may not cause a further increase. This appears to be unlikely since the spectral distribution of the afterglow, including the infrared region, is believed not to change as it decays at constant temperature and pressure. The time during which the decay occurs is long compared to the time of 1 or 2 milliseconds required for active nitrogen to travel through a supersonic nozzle or other aerodynamic test section of flow. If the half-life of the afterglow is 1 second, a temperature rise of about 0.05°C may be expected in a stream tube during a time of 0.001 second required for an element of the active gas to move a distance of the order of 2 feet at supersonic velocities. A temperature rise of this amount is considered to be acceptable in aerodynamic applications of the afterglow.

Grateful acknowledgement is made to Dr. John D. Trimmer of the University of Tennessee, Department of Physics, for helpful discussions and suggestions during the investigation. Acknowledgment is also made to Dr. Joseph Kaplan of the University of California, Los Angeles, California, who first suggested that the Lewis-Rayleigh afterglow might be useful in the visualization of low density flow.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Experimental Verification of Source Size Theory for the Mach-Zehnder Interferometer

G. D. KAHL AND F. D. BENNETT
Ballistic Research Laboratories, Aberdeen Proving Ground, Maryland
(Received March 3, 1952)

The simplified theory of the Mach-Zehnder interferometer developed by the latter author is tested. The reduction in the visibility of fringes formed by the interferometer due to an extended area light source is shown experimentally. This reduction is shown to be caused by the superposition of two separate effects which are characteristic of the two angular coordinates of the unit vector pointing from the center of the collimating lens toward each point in the plane of the area-type source. The relation of each of the effects to that coordinate characterizing it is demonstrated experimentally to agree with the theory. The result of combining these two effects from all of the points of an area-type source is experimentally seen to cause the "focusing" of the visible fringes and to limit them to a finite region of the intersecting beams determined by the size and shape of the source.

1. INTRODUCTION

IT is desirable for certain uses of the Mach-Zehnder interferometer to have clear fringes of high intensity. The possibility of using an extended area type source of nearly monochromatic light has been investigated by E. H. Winkler[1] and F. D. Bennett.[2] The latter author developed a concise vector treatment relating the type of source and the clarity of the fringes. This vector treatment of the interferometer, extended in a subsequent paper,[3] implies several effects which may be tested experimentally and which give a clear picture of the relation between light source and fringe clarity. It is with these experimental tests that the present paper is concerned.

2. PATH DIFFERENCE FORMULA. NOTATION

The interferometer forms fringes by producing two coherent virtual light sources from one real source. If the plates of the instrument are specified by their unit vector normals (Fig. 1), the two coherent light sources correspond to rays reflected from mirrors $\bar{n}_2$ and $\bar{n}_4$. It is assumed that an area type monochromatic light source is made up of incoherent point sources; from each point the collimating lens produces a beam of parallel light which may be designated by the unit vector $\bar{r}$ directed from the lens center to the point. The reflection of this beam from an element is identified by a subscript corresponding to that of the reflecting element. Thus, beam vectors $\bar{r}_2$ and $\bar{r}_4$ identify the two

[1] E. H. Winkler, N.O.L. Report No. 1077, December, 1947; N.O.L. Report No. 1099, February 1950.
[2] F. D. Bennett, J. Appl. Phys. **22**, 184 (1951); BRL Rep't No. 731 (September, 1950).
[3] F. D. Bennett and G. D. Kahl, *Theory of Mach-Zehnder Interferometer* (to be published).

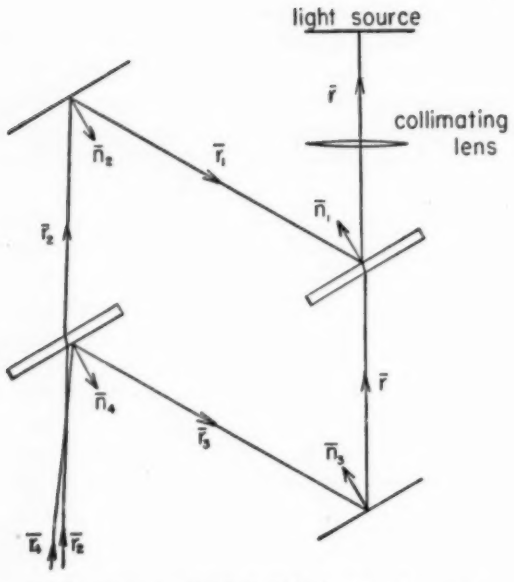


FIG. 1. Diagram of Mach-Zehnder interferometer (light rays reversed for convenience).

coherent beams produced by beam $\bar{r}$ after final reflection from plates $\bar{n}_2$ and $\bar{n}_4$.

In a polar coordinate system with origin at the lens center, the source points may be expressed in terms of the azimuth and elevation angles (α and β) of the beam vectors. Now we propose to show that in such a reference system, oriented with one axis parallel to the clear fringe direction,* the azimuth and elevation angles, respectively, will be uniquely associated with two different effects which limit the area of sharp fringes on any plane of observation.

It is assumed that the interferometer is adjusted so that one of the virtual light sources appears to be displaced from the other by a pure rotation about some fixed axis and that the compensation for the glass plates is perfect. Under these conditions† the beams $\bar{r}_2$ and $\bar{r}_4$ produced by division of beam $\bar{r}$ interfere at an arbitrary field point with a path difference of

$$d=(\bar{r}_2-\bar{r}_4)\cdot\bar{x}, \tag{1}$$

where $\bar{x}$ originates from the fixed rotation axis and identifies the field point. For any undivided beam $\bar{r}$ the vector $(\bar{r}_2-\bar{r}_4)$ is always perpendicular to the axis of rotation, since unit vectors $\bar{r}_2$ and $\bar{r}_4$ make the same angle with this axis; hence, $\bar{x}$ may be always taken in a plane perpendicular to this axis with no change in the value of d.

Equation (1) is recognized as the equation of a family of parallel planes with normal $(\bar{r}_2-\bar{r}_4)$. The plane corresponding to a path difference $d=n\lambda$ is one of constructive interference. Each source point produces a family of such planes which appear as fringes when traced on an observation plane, e.g., a photographic plate. The field consists of a superposition of families of fringe planes.

3. COORDINATE SYSTEM

Defining a unit orthogonal triad $\bar{\imath}, \bar{\jmath}, \bar{k}$ (Fig. 2) with $\bar{\imath}$ taken along the rotation axis, i.e., along the clear fringe direction, the vectors $\bar{r}_2$ and $\bar{r}_4$ are expressed as

$$\bar{r}_2=\bar{\imath}\sin\beta+\bar{\jmath}\cos(\alpha+\epsilon/2)\cos\beta+\bar{k}\sin(\alpha+\epsilon/2)\cos\beta,$$

$$\bar{r}_4=\bar{\imath}\sin\beta+\bar{\jmath}\cos(\alpha-\epsilon/2)\cos\beta+\bar{k}\sin(\alpha-\epsilon/2)\cos\beta,$$

and it is clear that ϵ is the amount of rotation about the axis that one image is displaced from the other.

A similar triad translated without rotation to the center of the collimating lens may be taken as the orientation of the particular coordinate system used to identify the real source points.

From the above, vector $(\bar{r}_2-\bar{r}_4)$ becomes

$$(\bar{r}_2-\bar{r}_4)=2\cos\beta\sin\epsilon/2(\bar{k}\cos\alpha-\bar{\jmath}\sin\alpha), \tag{2}$$

and as angle α changes, vector normal $(\bar{r}_2-\bar{r}_4)$ appears to rotate about the $\bar{\imath}$ axis of rotation.

By taking $\bar{x}$ parallel to $\bar{r}_2-\bar{r}_4$, $(\bar{x}=x\{\bar{k}\cos\alpha-\bar{\jmath}\sin\alpha\})$, then $d_2-d_1=\lambda=2\sin\epsilon/2\cos\beta(x_2-x_1)$, and $S=x_2-x_1$ is the perpendicular spacing between the planes of this family:

$$S=\lambda/(2\sin\epsilon/2\cos\beta). \tag{3}$$

Physically, $\bar{x}$ is taken to lie in a photographic plate and therefore is fixed rather than free to rotate with the normal to each of the families of interfering planes

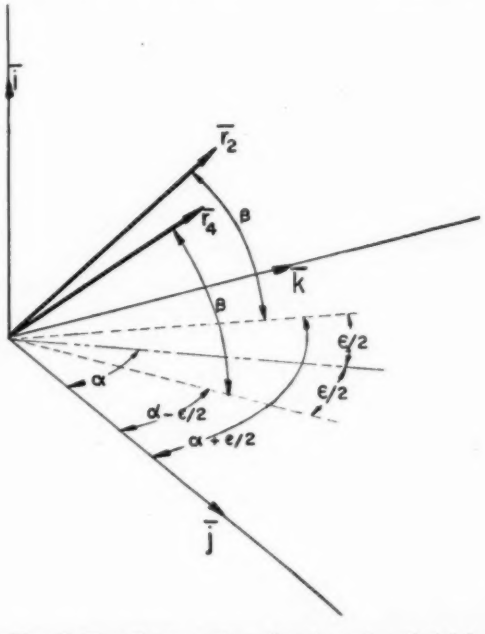


FIG. 2. Coordinate system showing azimuthal (α) and elevation (β) angles.

* This direction depends on the choice of plates moved to produce fringes. If only the $\bar{n}_4$ vector is moved from the zeroth fringe position, the clear fringe direction is given by $(\bar{n}_4\times\bar{n}_2)$ and may have any direction in space consonant with small rotations of $\bar{n}_4$ (v. reference 2, Secs. IV and V).

† For a more general form of the path difference formula which includes thick plate compensation and a more general type of image displacement, see reference 3.

occurring for each $\bar{r}$ vector. It is usually convenient to fix the photographic plate parallel to the normal of the family of planes formed by the central source point characterized by unit vector $\bar{r}_c$. Then $\bar{x}$ is parallel to $(\bar{r}_{2c}-\bar{r}_{4c})$ and

$$\bar{x}=x(\bar{k}\cos\alpha_c-\bar{j}\sin\alpha_c);$$

thus,

$$d_4-d_3=\lambda=2\sin\epsilon/2\cos\beta\cos(\alpha-\alpha_c)(x_4-x_3).$$

Here (x_4-x_3) is the fringe width appearing on the photographic plate which differs from the perpendicular spacing between planes of a family by the factor $1/[\cos(\alpha-\alpha_c)]$:

$$w=x_4-x_3=\frac{\lambda}{2\sin\epsilon/2\cos\beta\cos(\alpha-\alpha_c)}=\frac{S(\beta)}{\cos(\alpha-\alpha_c)}. \quad (4)$$

Equations (2) and (3) show that the azimuth angle α of a source point fixes the inclination of the family of fringe planes, while the elevation angle β determines the spacing between planes of this family.

4. EXPERIMENTAL TECHNIQUE

The experimental technique consisted of exploring the field with test planes (photographic plates) for different types of sources. Because the vector $(\bar{r}_2-\bar{r}_4)$ is always perpendicular to the $\bar{\imath}$ axis, every family of planes is perpendicular to the $\bar{j}$, $\bar{k}$ plane. Traces in the $\bar{j}$, $\bar{k}$ plane formed by the family corresponding to a single source point may be examined for inclination and spacing, i.e., for the effects of α and β. Traces formed by fringe planes from more than one source point superpose and show the effects on contrast of the composite fringes of the (α, β)'s corresponding to the separate points. All pictures shown were obtained from plates oriented in or within about 30° of this plane; thus a "cross section" of the fringe region is presented. When examining the pictures, it is helpful to think of the direction along the fringes as that corresponding to looking directly into the interferometer.

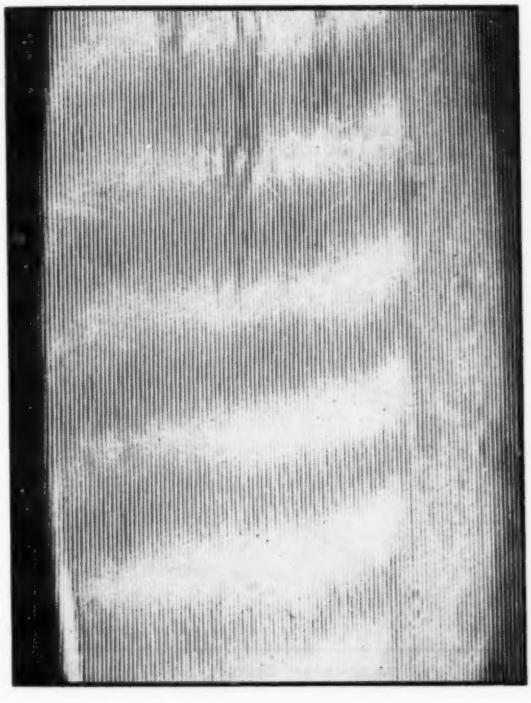


FIG. 3. Fringes from two-point source with β=const.

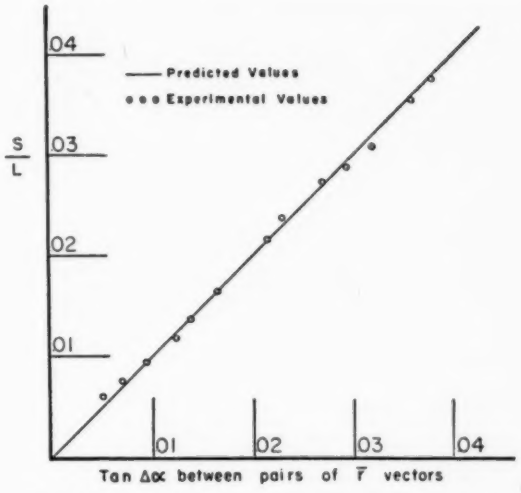


FIG. 4. Verification of "fade-out" dependence on α.

The interferometer used for the tests had 2-in. diameter plates centered on a plane with the light incident at 30° on the first divider. Real fringes were localized outside the interferometer to avoid the use of a camera lens system.

Figure 3 shows the results of one photograph taken with the fringes oriented perpendicular to the plane of centers, corresponding to $\beta_c=0$. A two-point source was used with both source points having a constant elevation angle so that the spacing in each family is equal. The families are inclined, however, because of the difference in the azimuth angles α_1 and α_2. At the edges of the picture, light from one source point was cut off by the aperture, thus accounting for the relatively clear fringes from the other point. If "l" is the distance between the periodic blurring, $S/l \doteq \tan(\alpha_2-\alpha_1)$. Figure 4 verifies that S/l approximately equals the tangent of the difference of the azimuth angles of such pairs of source points. These data were collected from a group of pictures similar to Fig. 3.

If many point sources of constant angle β are admitted, the clearest fringes will be seen only in a region near the rotation axis, since along this axis $\bar{x}=0$, and the path difference is constant $(d=0)$ for all source points. This localizes the fringes to a flat cylindrical region with generators parallel to $\bar{\imath}$, seen in cross section in Fig. 5. The superposition of a number of inclined fringe families destroys the contrast elsewhere. Ordinarily these fringes would be observed on an observation plane perpendicular to the plane of Fig. 5; for sharp

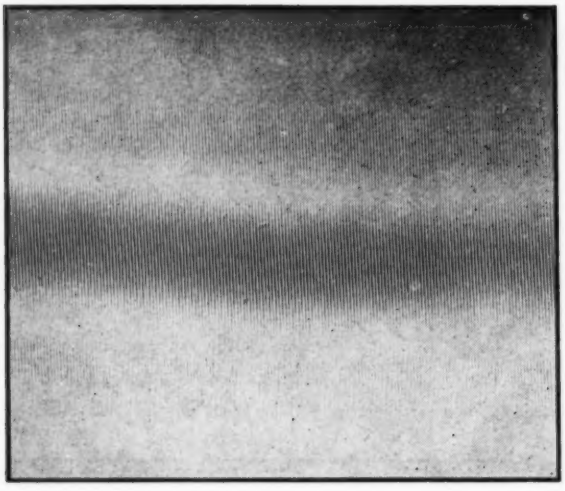


FIG. 5. Line source with β=const showing depth of fringe region.

fringes the position of this observation plane is critical owing to the small depth of "focus" caused by the extended source.

The remaining pictures were taken with the fringes oriented parallel to the plane of centers, corresponding to $\beta_c=30°$. Figure 6 shows two superposed families caused by two source points with constant azimuth angle. The two families are therefore parallel, but their unequal spacing (due to the difference of the β-angles) causes periodic blurring parallel to the fringe planes. The faded areas occur where an interfering plane of one family lies midway between two planes of the other. The observed distance between fade outs is $l=NS_1=(N+1)S_2$ (N being an integer), and since $N(S_1-S_2)=S_2=N\Delta S$ and $S_2\doteq S_1$, we have

$$\Delta S/S\doteq S/l, \tag{5}$$

where S is the spacing in one family and ΔS is the difference in the spacing of the two families. From Eq. (3)

$$dS/S=\tan\beta_c(d\beta) \quad \text{or} \quad \Delta S/S=\tan\beta_c(\Delta\beta)\doteq S/l. \tag{6}$$

This equation holds for points having elevation angles in the neighborhood of β_c. If β_c is very small, corresponding to the "vertical" fringe orientation, l becomes large; for this reason values of $\beta_c\doteq 30°$ were used to demonstrate this type of fringe confusion.

Figure 7 shows the agreement between predicted and observed values of $\Delta S/S$ for a variation of angle β between pairs of source points. These data were collected from a group of pictures similar to Fig. 6.

If the source consists of a fairly large locus of points with constant azimuth angle, the fringes remain sharp only in a region centered on the rotation axis, as shown in Fig. 8. For this picture $\Delta\beta_{max}\doteq 2°$. Since these fringes would ordinarily be observed on a plane perpendicular to that of Fig. 8, the position of this plane for best

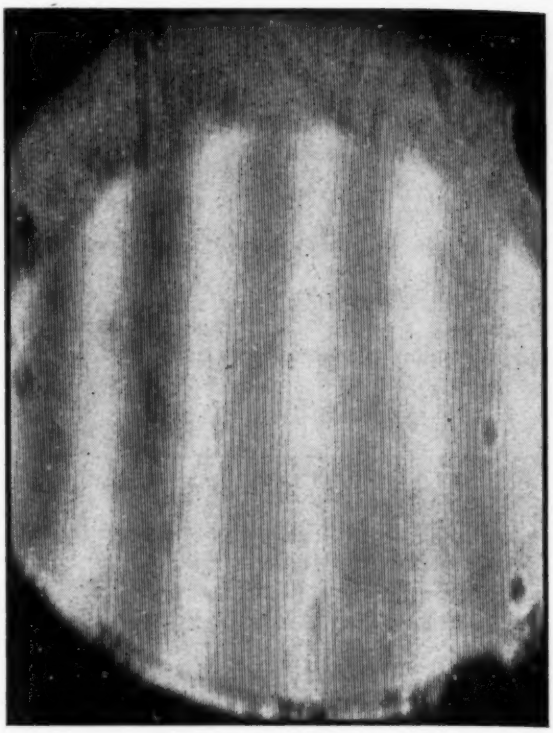


FIG. 6. Fringes from two-point source with α=const.

fringes is not critical, corresponding to a large depth of focus.

If an extended area light source is used, clearest fringes are confined to a small cylindrical region around the rotation axis, shown in cross section in Fig. 9. This picture was taken with fringes oriented parallel to the plane of centers ($\beta_c\doteq 30°$) using a circular source with a maximum angular spread of 2° from the lens center. The rotation axis is perpendicular to the plane of Fig. 9 in the center of clearest fringes. If Fig. 5 is superposed on Fig. 8, the common region of good fringes corresponds approximately to that found in Fig. 9.

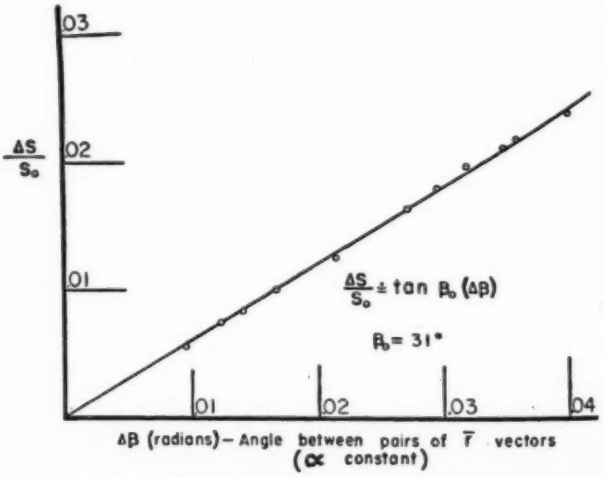


FIG. 7. Verification of "fade-out" dependence on β.

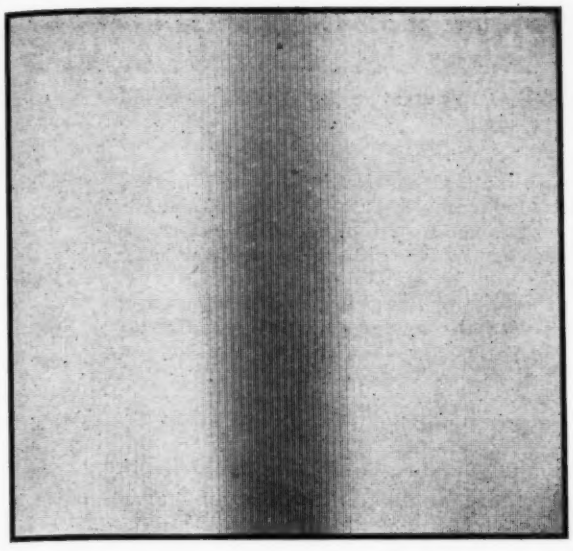


FIG. 8. Line source with α=const showing depth of fringe region.

5. PREFERRED SOURCE ORIENTATION

To obtain a given number of clear fringes in the field for a particular adjustment of the interferometer the angles α and β of the beam vectors must be restricted according to the limitations discussed in reference 2. This analysis shows that the optimum source may be used for a fringe orientation perpendicular to the plane of centers. The admissible source is then a circle of radius ρ,[4]

$$\rho=2f(K/N)^{\frac{1}{2}}, \tag{7}$$

where f is the focal length of the collimating lens, and N is the number of fringes formed by the $\bar{r}_c$ family on a photographic plate to which the fringes from all other families coincide to within $\pm K$ fringe widths. The photographic plate should be perpendicular to $(\bar{r}_{2c}+\bar{r}_{4c})$. If a slit source is to be used instead of the whole circle, greater depth of focus results if the long dimension of the slit is oriented parallel to the fringes.

[4] The area of this circle is twice as large as that reported in reference 2. Equation (7) may be derived by the methods of reference 2 by replacing $\bar{r}_d\cdot\bar{p}$ for $\bar{r}_c\cdot\bar{p}$ in Eq. (27) of reference 2, where $\bar{r}_d\cdot\bar{p}=1-(K/N)$. Equation (39) becomes $1-(2K/N)\leqslant\bar{r}\cdot\bar{p}\leqslant 1$ from which the above value of ρ may be found.

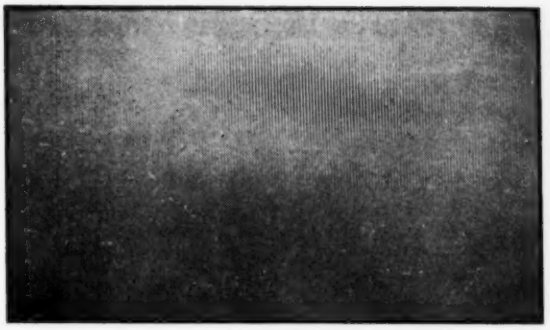


FIG. 9. Patch source α, β variable showing extent of fringe region.

6. ADJUSTMENT OF THE INTERFEROMETER

The path difference formula in Sec. 2 is a simplification of a more general expression derived in reference 3 and is valid only under the conditions stated in Sec. 2. In the general expression there is an additional term $\bar{q}$ which describes the displacement of one source image from the other by a translation as well as a rotation. This effect uniformly reduces the contrast of the fringes from any given source. By rotating one of the elements of the instrument to reduce the translation component of the image displacement, the fringes should increase in contrast until this component is zero. Experimentally, it is observed that in a multiple plate adjustment a slight rotation of one of the elements in a manner that rotates the orientation of the fringe increases (or decreases) markedly the fringe contrast. At the adjustment for maximum contrast this term $\bar{q}$ is assumed zero.

The Dielectric Properties of Ice and Snow at 3.2 Centimeters

W. A. CUMMING*
National Research Council, Ottawa, Canada
(Received February 14, 1952)

A knowledge of the permittivity and loss tangent of snow and ice is essential in studying both radar echoes from snow-covered terrain and the attenuation of microwave energy through snowstorms. To provide this information, a program was carried out at the National Research Council of Canada to measure the permittivity and loss tangent of ice and snow, and also to determine the reflection coefficients of snow-covered surfaces, at a wavelength of 3.2 centimeters.

As a result of these investigations values have been obtained for the permittivity and loss tangent of snow of varying density, temperature, and water content. Theoretical values of the reflection coefficients of snow-covered surfaces, calculated from these data, are compared with the values obtained from the direct measurement of reflections from natural snow surfaces.

I. OUTLINE OF INVESTIGATIONS

A STUDY has been made at the National Research Council of the dielectric properties of ice and snow and of the relation of these properties to the reflection coefficients of snow-covered terrain. Two separate investigations were carried out: one to determine the permittivity and loss tangent of ice and snow using wave-guide techniques; the other to determine by direct measurement the reflection coefficients of snow-covered surfaces. The results of the first investigation yielded information which can be applied to any problem concerning the electrical properties of snow, while the results of the second investigation can be used to predict the effect of a snow covering on the echo received from a radar target.

All the investigations described in this paper were carried out at a frequency of 9.375 kilomegacycles.

II. MEASUREMENT OF PERMITTIVITY

The permittivity of snow of varying density, and of ice, was obtained from the direct measurement of guide wavelength in a rectangular three-centimeter wave guide filled with snow or ice. A conventional wave-guide slotted section some eight inches in length was used, terminated with a short circuit in order to make the determination of the voltage minima as accurate as possible. Extreme care was exercised in filling the section with snow in an effort to maintain constant snow density throughout the wave guide; and in the case of ice, water was poured in the wave guide a few cubic centimeters at a time and allowed to freeze so that no distortion of the slotted section would occur. The electrical length of the slotted section varied between four and seven half-wavelengths, depending on snow density, so that for each density a mean value of guide wavelength was obtained, thus minimizing errors due to nonuniform snow density. In general, the variation of individual measured half-wavelengths was of the order of one percent of the mean value for each sample.

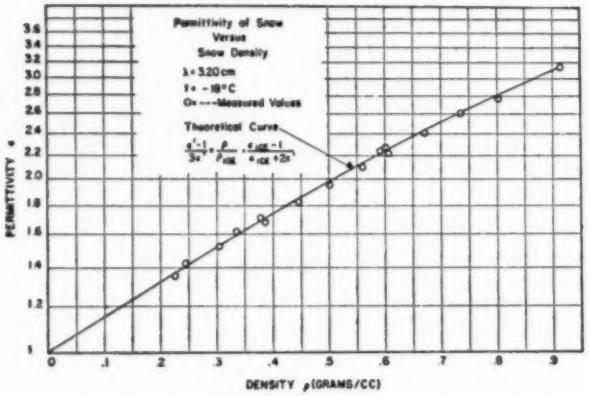


FIG. 1. The relation between snow permittivity and density.

At the beginning of the measurements a check was made to determine whether or not the permittivity was temperature-dependent in the range 0° to −20°C. Only random variations of permittivity were noted —of the order of ±0.5 percent—so that the permittivity can be said to be constant in this temperature range. This was to be expected since anomalous dispersion in ice occurs at a frequency of approximately 10^4 cps.

The measurements of guide wavelength as a function of snow density were accordingly carried out at a fixed temperature of −18°C. Permittivity was then calculated from guide wavelength, frequency, and wave-guide dimensions. In making these calculations the loss of the material filling the wave guide was not considered, since preliminary measurements had shown that the loss tangent of ice and snow at −18°C was too small to have a significant effect on guide wavelength. In the case of ice, the effect of the loss component is such as to produce a change in guide wavelength of less than 1/20th of one percent.

The experimental values of permittivity as a function of density are shown in Fig. 1. The value for ice was found to be 3.15, which is in close agreement with the value of 3.1_7 obtained by Lamb,[1] who used a cavity

* Assistant Research Officer, Radio and Electrical Engineering Division, National Research Council, Ottawa, Canada.

[1] J. Lamb, Proc. Phys. Soc. (London) **B62**, 272 (1949).

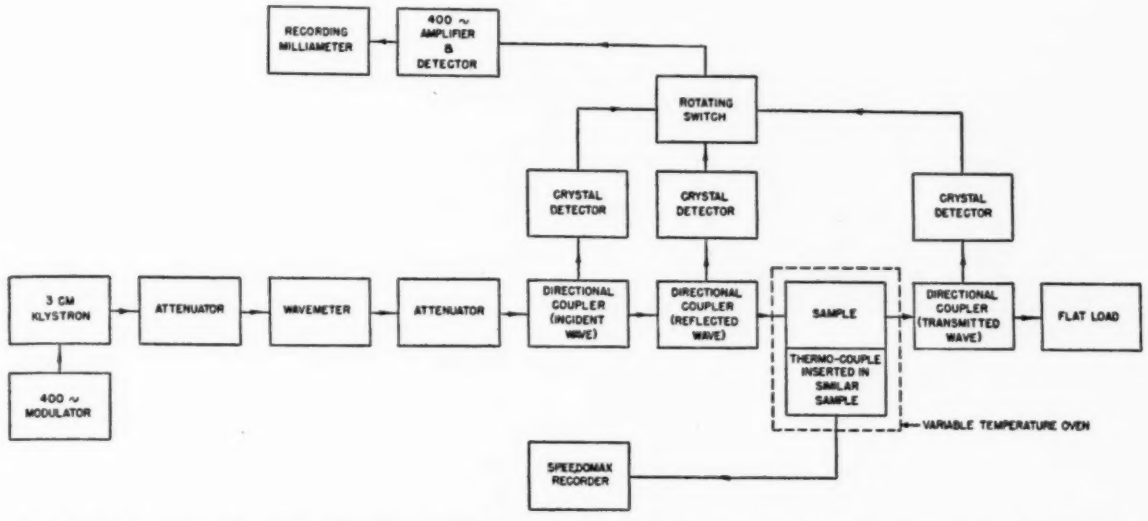


FIG. 2. Block diagram of automatic equipment used to measure the loss of snow and ice at three-centimeter wavelength.

method of measurement. The smooth curve drawn through the experimental points is a plot of a dielectric mixing formula

$$\frac{\epsilon'-1}{3\epsilon'}=f\frac{\epsilon_0-1}{\epsilon_0+2\epsilon'},$$

due to Polder and van Santen,[2] where ϵ_0 is the permittivity of a solid dielectric (in this case ice), f the volume ratio of the solid dielectric to the total volume of the mixture of the dielectric particles and air, and ϵ' is the permittivity of the mixture. This expression is a particular case of a general formula for ellipsoidal particles of the solid dielectric which simplifies to the above form when the particles are assumed to be spherical. In plotting the theoretical curve the measured value of 3.15 for the permittivity of ice was used. As can be seen, the curve fits the experimental points quite well, though it has the general tendency to be slightly higher than the measured values in range of density between 0.4 and 0.7 grams/cubic-centimeter.

III. MEASUREMENT OF LOSS TANGENT

The basic quantity measured to determine the loss tangent of ice and snow was the transmission through a section of three-centimeter wave guide filled with the material under test. This transmission was measured continuously as a function of temperature by using directional couplers whose outputs were automatically recorded. A block diagram of the apparatus is shown in Fig. 2. The sample holder—a length of wave guide approximately eight inches long—was contained in an insulted oven whose temperature could be cycled automatically between 0° and −18°C. The inside length of the oven was some fourteen inches, so that the temperature was uniform throughout the length of the sample.

Modulated three-centimeter energy was fed to the sample from a well-padded klystron oscillator mounted in a temperature-controlled oven. Two directional couplers were interposed between the oscillator and sample to give a measure of the amplitude of the incident and reflected traveling waves. A third directional coupler was located between the sample and a matched load which terminated the system. The three crystal detectors associated with the directional couplers were mounted in the temperature-controlled oven containing the oscillator. The outputs of these detectors were fed into a rotating switch and thence to an amplifier and recording milliammeter. The switch speed was such that each signal was recorded for an interval of forty seconds for each rotation of the switch. Thus a semi-continuous record of the incident, reflected, and transmitted waves was obtained.

The temperature of the oven containing the sample was measured by means of thermocouples and a recording meter. The thermocouples were contained in a snow- or ice-filled section of guide identical with the sample holder and mounted beside it. Thus a record of the actual temperature of the snow or ice under test was obtained. In order to assure uniform cooling of the snow, the oven temperature was varied slowly, an interval of one and a half hours being required for the temperature to change from 0° to −18°C, the temperature of the cold chamber housing the apparatus.

Extreme care was exercised in order to obtain identical characteristics in the three directional couplers. The final adjustment was made by setting each coupler to record a given wave, and then adjusting the attenuators between the couplers and their detectors so that each detector gave the same reading. In order to check the accuracy of the equipment, the loss tangents of several materials of known loss were measured and were found to be in good agreement with the values

[2] D. Polder and J. H. van Santen, Physica 12, 257–271 (1946).

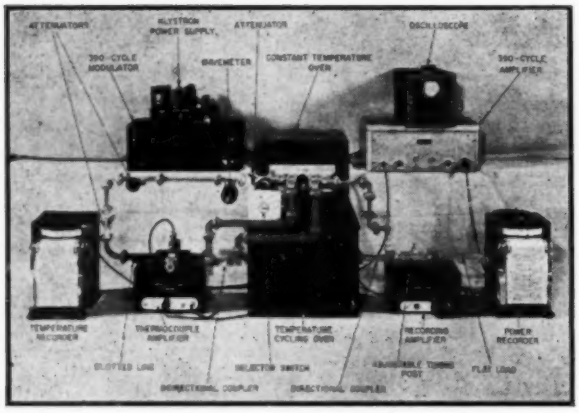


FIG. 3. Automatic equipment for the measurement of dielectric loss in snow and ice at three-centimeter wavelength.

obtained by other observers. A photograph of the complete apparatus is shown in Fig. 3.

As in the case of the permittivity measurements, the snow and ice samples were made as homogeneous as possible. In order to check the effect of the dielectric windows on each end of the sample holder, measurements were made on solid dielectrics such as Teflon, with and without the windows, and showed no significant differences. Errors due to finite discrimination of the directional couplers were minimized by making the sample of such a length as to give a minimum reflected wave. In addition, every attempt was made to eliminate water condensation in the wave-guide sections adjacent to the sample holder by having these sections completely sealed off. As a check on the effectiveness of these measures, the loss tangent of Teflon was measured both at room temperature and in the cold chamber, with no significant difference in results.

In operation then, the directional couplers were calibrated, the sample inserted, and the temperature cycled from -18°C up to 0°C two or three times, after which the calibration was rechecked. The loss tangent was then calculated as a function of temperature from the coupler readings, frequency, wave-guide dimensions, and the permittivity, as obtained earlier. The total measured attenuation was of course corrected to allow for the wall loss of the sample holder, and thus the true attenuation of the snow was used in computing the loss tangent. This was done for several snow samples of varying density.

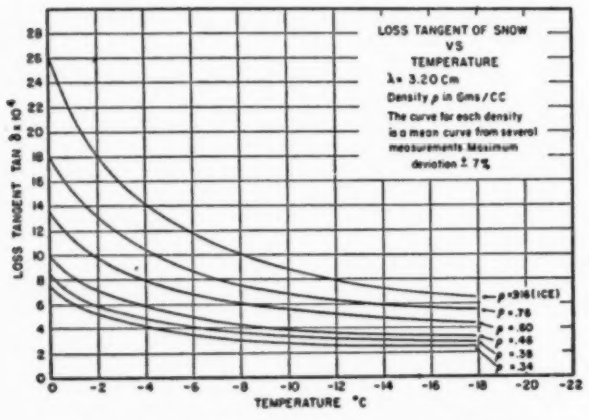


FIG. 4. The variation of the loss tangent of snow with temperature.

The loss tangent of ice as a function of temperature is shown in Fig. 4. Since the transmission was recorded continuously, no individual points are shown. This curve is a mean of four curves, the maximum deviation being approximately -5 percent. As can be seen, the ice behaves as a reasonably low loss dielectric in this range, with a maximum loss tangent of 27.0×10^{-4}. The loss of ice at three centimeters has also been measured by Lamb[3] and the M.I.T. Dielectrics Laboratory,[4] whose values can be compared. The M.I.T. measurement was carried out at a fixed temperature of -12°C, and gave a value of $\tan\delta=7\times10^{-4}$. The corresponding value as obtained here was 7.8×10^{-4}. However, the agreement with Lamb's work is very poor, the values being roughly twice his. It has been suggested that the nature of the crystal growth might be different for the two cases, and that this could have an effect on the loss tangent. No detectable difference was observed in the loss tangent of different samples made from distilled water, tap water, and melted snow.

The loss tangent of snow of varying density as a function of temperature is also shown in Fig. 4. As in the case of ice, these are mean curves, the maximum deviation being of the order of seven percent. In one or two cases the loss tangent appeared to have a permanent increase after the snow had been taken to 0° and returned to a lower temperature, but no particular significance is attached to this as it did not occur regularly. No correlation could be obtained between loss tangent and snow crystal structure.

In studying the practical problem of the attenuation through a snow layer, the increase in loss due to the presence of free water in the snow must be considered. In order to obtain an estimate of the effect of free water, two snow samples were allowed to remain at zero degrees until the transmission was almost zero. The free water content was then measured, and on the assumption that this free water content had increased linearly with time from the melting point, curves of loss tangent *versus* water concentration were obtained. These are shown in Fig. 5 and as can be seen the presence of only 0.8 percent free water is sufficient to increase the loss tangent by a factor of ten.

IV. MEASUREMENT OF REFLECTION COEFFICIENT

The magnitude of the reflection coefficient at three centimeters was obtained by measuring the vertical

[3] J. Lamb, Trans. Faraday Soc. **52A**, 238–244 (1946).
[4] A. von Hippel, *Tables of Dielectric Materials* (Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts, 1945), Vol. II, p. 220.

radiation pattern of an antenna mounted at varying heights above the surface under study. This is shown in the sketch of Fig. 6. Two three-centimeter slotted wave-guide arrays were mounted horizontally on a carriage which moved vertically to a height of thirty feet. These arrays were designed to have narrow beam widths in the horizontal plane (of the order of 5°) and extremely broad beams in the vertical plane, and were polarized normal to the plane of incidence. The spacing between the arrays was such that the reflection coefficient could be measured at grazing angles up to 50°. The ground between the towers was marked off in an area 40 feet by 12 feet—large enough to contain the first Fresnel ellipse. The area, consisting of very sandy soil, was made as smooth as possible in order to give true specular reflection. This was accomplished by placing 40-foot planks on either side of the prepared area, the planks serving as rails over which a metal

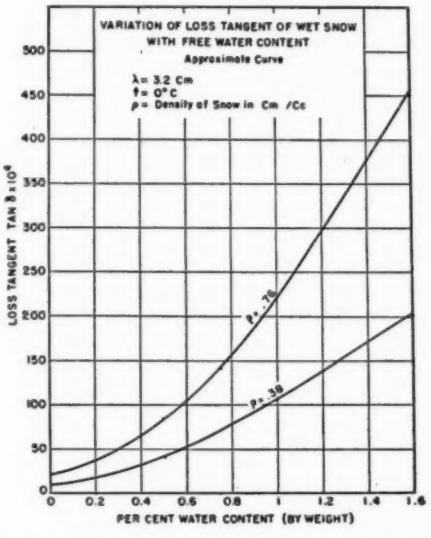

VARIATION OF LOSS TANGENT OF WET SNOW
WITH FREE WATER CONTENT
Approximate Curve
λ= 3.2 Cm
t= 0°C
ρ= Density of Snow in Gm /Cc
LOSS TANGENT TAN δ x 10⁴
PER CENT WATER CONTENT (BY WEIGHT)
ρ= .7%
ρ= .3%

FIG. 5. The effect of free water on the loss tangent of wet snow.

scraper could be pulled to level off the soil or snow between the planks.

In operation then, one of the antennas—say the transmitter—was fixed at a given height above ground. Owing to the fact that the vertical pattern of the antenna was not truly circular, it was necessary that the antenna be tilted so that the line of the main beam bisected the angle between the direct and reflected paths between the two antennas (see Fig. 6). The receiving array was then tilted in the same fashion and allowed to move slowly in a region extending one foot above and below the height of the transmitting antenna. As the antenna moved, the received signal was plotted automatically as a function of height, so that a continuous record of the interference pattern was made in the region. The transmitter was then moved to a point two feet higher and the same procedure carried out. In this manner a continuous plot of the interference

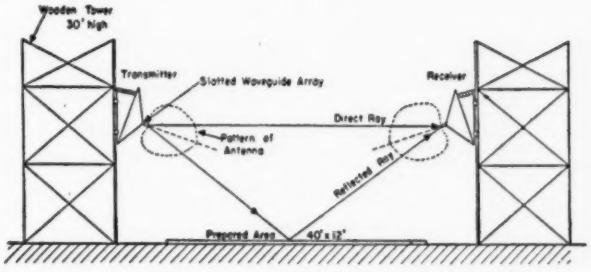


FIG. 6. Diagram of apparatus used to measure reflection coefficients.

pattern was obtained, from which the reflection coefficient could be calculated.

The first measurement made with this apparatus was of the reflection coefficient of a flat frozen sand surface. Prior to freezing the sand had been slightly moist so that a flat frozen crust of sand constituted the reflecting surface. The measured values obtained over this surface are shown in Fig. 7. In its earlier form the apparatus did not record continuously, so that a number of discrete values of reflection coefficient were obtained, as shown on the graph.

It was found that the best-fit theoretical curve drawn through the points corresponded to a dielectric constant of 2.1 for the reflecting medium. This best-fit curve is shown in Fig. 7. The value of 2.1 is in excellent agreement with the value of 2 obtained by Ford and Oliver[5] over dry sand at nine centimeters, and in fact the slight increase could be due to the small amount of ice interspersed through the sand.

Two typical curves for the reflection coefficient of this sand surface when covered with snow are shown in Figs. 8 and 9. The snow of Fig. 8 was measured shortly after it fell, and made a covering of some nine inches of very slightly moist snow (water content less than $\frac{1}{4}$ of 1 percent). No attempt was made to level

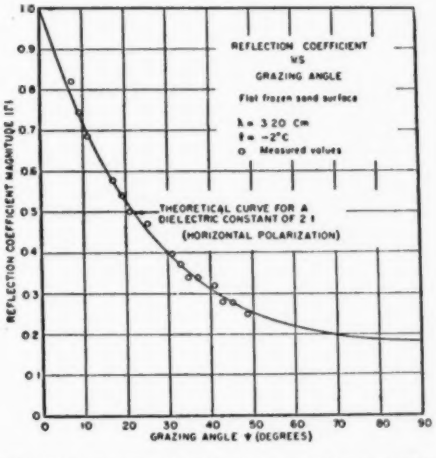


FIG. 7. The measured reflection coefficient of a frozen sand surface at 3 cm.

[5] L. H. F. Ford and R. Oliver, "An experimental investigation of the reflection and absorption of radiation of 9-cm wavelength," Radio Division, National Physical Laboratory, 1944.

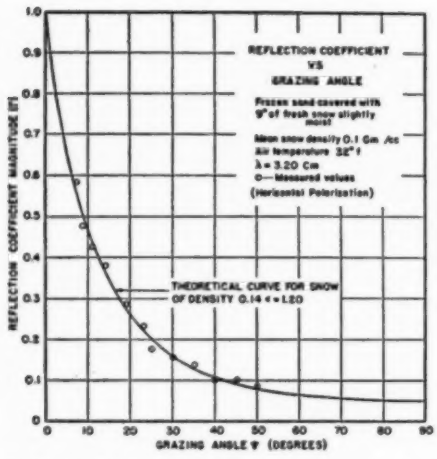


FIG. 8. The reflection coefficient of a frozen sand surface covered with moist snow.

the snow, as it fell in the absence of wind, and hence gave quite a level surface. As in the case of sand, a best-fit curve was drawn, which corresponded to a dielectric constant of 1.2, and hence a density of 0.14 gram/cubic-centimeter. The measured mean snow density was 0.1 gram/cubic-centimeter. There was no evidence of multiple reflections in this measurement, and the best-fit curve shows all the reflection occurred at the air-snow surface.

The curve of Fig. 9 was measured over granular snow which appeared to be quite dry. This measurement was carried out at an air-temperature of 31°F, and again no indications of multiple reflections were present. The mean measured snow density was 0.25 and the density for a best-fit curve was 0.21; again a reasonable agreement.

These two curves are quite typical of the measurements obtained with the flat sand surface covered with snow. In no case were multiple reflections evident; all indications pointing to the fact that the reflection at the air-snow boundary was predominant.

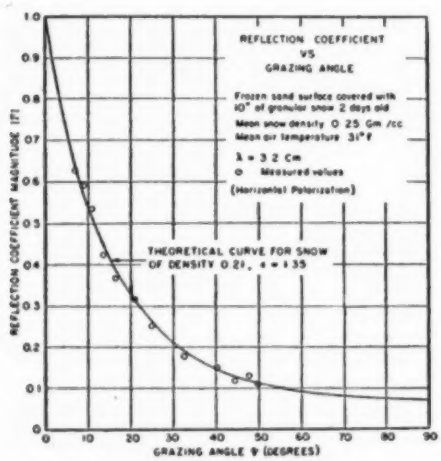


FIG. 9. The reflection coefficient of a frozen sand surface covered with 10 in. of dry snow.

In an effort to learn more about the attenuation of an natural snow covering, the prepared sand area was replaced with a metal platform, made of sheet aluminum. Measurements were then made of the reflection coefficients of this surface covered with various snow blankets.

A typical curve taken over moist snow is shown in Fig. 10. Here again, in spite of the excellent reflecting surface beneath the snow, no multiple reflections were present. In this measurement a continuous recording of reflection coefficient was obtained, so that the theoretical curve drawn for comparison is one based on the actual mean measured snow density. The measured curve of reflection coefficient over the metal surface is also given for comparison. This curve is quite typical of measurements made over moist snow, showing that even when an almost perfect reflector lies beneath the snow, the predominant reflection is from the air-snow boundary.

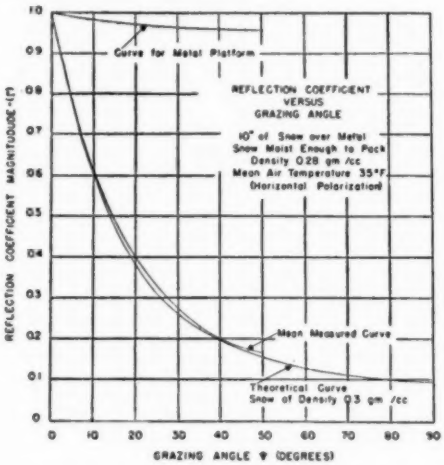


FIG. 10. The reflection coefficient of a metallic surface covered with moist snow.

Figure 11 and 12 show the type of reflection that occurred when the metal platform was covered with dry snow, well below the freezing point. Multiple reflections were quite evident in these cases, particularly for thin coverings of snow. Owing to the complex manner in which the reflection coefficient varied with grazing angle, only the envelope of the curve is drawn. In each case a theoretical curve for an infinite depth of snow of the measured density is drawn for comparison, along with the curve for the metal platform. As expected, the effect of the multiple reflections increases with grazing angle, owing to the shorter path through the snow and the larger transmission coefficient at the air-snow boundary. However, in spite of the low loss through the dry snow, the reflection coefficient was greatly reduced from its value in the absence of snow.

V. DISCUSSION OF RESULTS

A comparison of the measured values of loss tangent for snow and the measured values of reflection coeffi-

cient shows quite good agreement in the case of wet snow. Experiment showed the loss tangent of wet snow to be quite high, and this observation was borne out by the absence of multiple reflections over wet snow, even when a near-perfect reflector lay beneath the snow surface. In the case of dry snow, however, the measured values of loss tangent would predict a more noticeable effect due to multiple reflections than was noted. It may be possible that because of the non-uniform density of the snow covering scattering occurs within the snow, effectively reducing the reflection from the boundary between the snow and the surface beneath it.

In the cases where multiple-reflections were not present, excellent agreement was obtained between measured values of permittivity and measured values of reflection coefficient, particularly when the snow lay on a sand surface with its effective dielectric constant of 2.1.

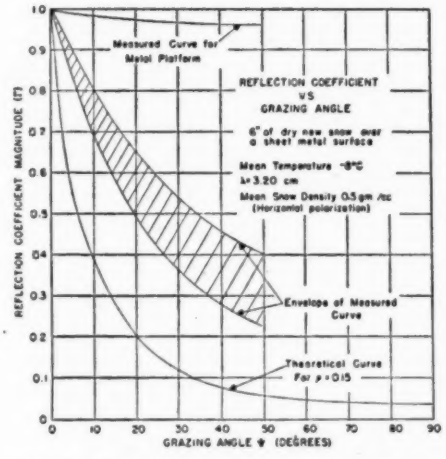

Measured Curve for Metal Platform
REFLECTION COEFFICIENT VS GRAZING ANGLE
6" of dry new snow over a sheet metal surface
Mean Temperature -8°C
λ=3.20 cm
Mean Snow Density 0.5 gm /cc (Horizontal polarization)
Envelope of Measured Curve
Theoretical Curve For p=0.15
REFLECTION COEFFICIENT MAGNITUDE (F)
GRAZING ANGLE ψ (DEGREES)

FIG. 11. The reflection coefficient of a metallic surface covered with 6 in. of dry snow.

Ford and Oliver[5] found in their extensive measurements at nine centimeters that over most terrain specular reflection would be unusual except at near zero grazing angle, since it can only occur at very level surfaces free of vegetation. This effect would be even more pronounced at the shorter wavelength of three centimeters. They estimate the probable reflection for rough ground to be 0.1, except near grazing incidence. Even over flat surfaces, such as sand, the reflection coefficient at normal incidence is only 0.2.

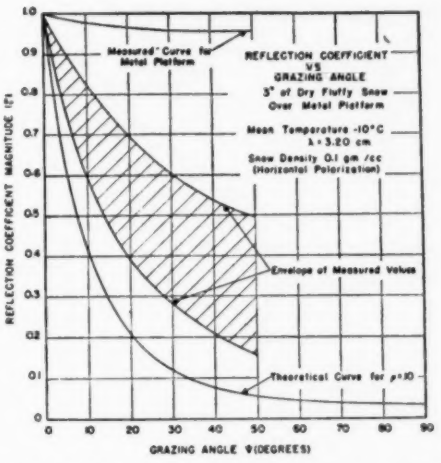

Measured Curve for Metal Platform
REFLECTION COEFFICIENT VS GRAZING ANGLE
3" of Dry Fluffy Snow Over Metal Platform
Mean Temperature -10°C
λ=3.20 cm
Snow Density 0.1 gm /cc (Horizontal Polarization)
Envelope of Measured Values
Theoretical Curve for p=10
REFLECTION COEFFICIENT MAGNITUDE (F)
GRAZING ANGLE ψ (DEGREES)

FIG. 12. The reflection coefficient of a metallic surface covered with 3 in. of fluffy snow.

In applying these results to radar performance over snow-covered terrain, it may be said that at small grazing angles an echo would usually be reduced due to the snow covering. In general, a radar echo is caused by both specular reflection and back scattering; but at near-zero grazing angles a specular reflection directs energy away from the antenna so that the return signal is due to back scattering. Owing to the smooth nature of a snow covering relative to the surface of the terrain, specular reflection is more apt to occur, with a consequent reduction in the echo at small grazing angles.

On the other hand, near normal incidence both scattering and reflection contribute to the radar echo. Except in the case of thin, dry, snow coverings over very good reflectors, the specular reflection is usually increased by the presence of a snow layer. This would be particularly true of rough terrain covered by either moist snow or a thick layer of dry snow, since with both types of snow the reflection coefficient at the air-snow surface would be of the order of 0.1 to 0.2 near normal incidence, while the reflection coefficient of the bare terrain would be less than 0.1.

ACKNOWLEDGMENT

The author wishes to express his thanks to Dr. G. A. Miller of the Radio and Electrical Engineering Division for his guidance in this project, and to Mr. E. P. Deloli, who assisted with the design of the apparatus and obtained the experimental data in the cold chamber.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Axial Motion of an Electron in a Constant Wave Velocity Section of a Linear Accelerator*

DAVID CAPLAN AND EDWARD AKELEY
Department of Physics, Purdue University, West Lafayette, Indiana
(Received March 31, 1952)

The relativistic differential equations for the axial motion of an electron in a wave guide, excited by a progressive, sinusoidal, longitudinal, electric wave of constant speed less than that of light in vacuum, have been integrated. The first integral yields the results of J. C. Slater and gives the energy-phase relationship, which is periodic. A second integration yields the phase distance relationship and thus completes the determination of the orbits. All separatrices are exactly integrable in terms of trigonometric and hyperbolic functions, while the general case requires elliptic integrals of the first and third kinds. The former are obtainable from tables, while the latter are evaluated by series expansions, or in the case of particles near the self-crossing point of the separatrices where the convergence of the series expansion is slow, by the use of the addition theorem for elliptic integrals of the third kind.

These results are applied to the case of the proposed new Purdue Linear Accelerator. Energy distribution of the particles coming out of the end of the first two sections of this accelerator is in this way obtained.

INTRODUCTION

AN electron on the axis of a constant velocity section of a traveling-wave guide experiences the accelerating action of the longitudinal, sinusoidal, axial electric field moving along the wave guide with constant phase velocity in a given section. It is the purpose of the present work to integrate the equations of motion of the electron in this type of field, for the case of the wave speed less than c, the speed of light in vacuum. (The case of wave speed equal to c will not be considered here.) From this integration, the orbits of the electron will be found, that is, the relationship of the energy of the electron, its phase relative to the wave, and the distance it has traveled along the wave guide. These results will be applied to the proposed new Purdue Linear Accelerator, the first two sections of which will have average maximum field strengths, 33 and 25 kilovolts per centimeter, and wave velocities $0.7c$ and $0.96c$, respectively. The sections are to be 12.985 cm and 104.34 cm long, respectively; all particles are to be injected into the first section with 50-kev kinetic energy.

These calculations are useful for determining the lengths and wave velocities of the wave-guide sections, bearing in mind the desirability of keeping as many injected electrons as possible in a region of maximum axial accelerating field and minimum radial defocusing field, which are 90° out of phase with each other. The results will also predict the energy distribution of the electrons coming out the other end of the accelerator.

The electric field strengths considered are not higher than 50 kilovolts per centimeter, so that it takes the electron at least 10 centimeters under the influence of this field before it acquires an added energy equal to its rest energy of about 500 kev (in the optimum case never actually realized in a constant velocity wave). It can be shown from classical considerations[1] that if this distance is of an order much greater than $e^2/m_0c^2=10^{-12}$ cm, then the self-radiation effects of the electron may be neglected in the equations of motion.

The slow variation (a few percent) of the maximum electric field strength with distance, owing to the attenuation of the wave guide, will be neglected. The higher Fourier components of the wave, which are relatively small and relatively swiftly moving by the electron, will also be neglected.

INTEGRATION OF THE EQUATIONS OF MOTION

We denote the total energy of the electron by U, the distance along the wave guide by X, the time by T, the angular frequency by ω, the charge on the electron by e, the maximum electric field strength by E, the constant speed of the wave by V, the rest mass of the electron by m_0, and the speed of light in vacuum by c. The velocity of the electron down the accelerator is thus given by dX/dT. To simplify the work, we introduce the following dimensionless quantities to represent the variables $u=U/m_0c^2$; $t=\omega T$; $x=(\omega/c)X=\lambda/2\pi$; $v=V/c$; $A=eE/m_0\omega c$; and $y=(x/v)-t$; the latter being the phase of the electron relative to the wave.

From special relativity theory, we obtain the equations of motion as follows:

$$du/dx=A\cos y, \tag{1}$$

$$dx/dt=(u^2-1)^{\frac{1}{2}}/u. \tag{2*}$$

The positive sign of the square root in Eq. (2*) has been taken because this problem involves only positive velocities (see later on). Thus,

$$dy/dx=(1/v)-u/(u^2-1)^{\frac{1}{2}}. \tag{2}$$

From Eqs. (1) and (2), by elimination of dx, we find by integration

$$h=A\sin y+Ak, \tag{3}$$

$$h=(u/v)-(u^2-1)^{\frac{1}{2}}. \tag{4}$$

* Presented at the Mexico City meeting of the American Physical Society. For abstract see Phys. Rev. **80**, 124 (1950), paper D3. A report on Electron Accelerators by E. L. Chu (M. L. Report No. 140, May, 1951, Microwave Laboratory, Stanford University) also treats the subject of this paper.

[1] J. Schwinger, Phys. Rev. **75**, 1912 (1949).

Wherein h is a variable depending upon total energy, which has been introduced for convenience, and k is an integration constant determined by the initial injection energy and phase of the electron. Thus, we now have

$$\cos y=\pm\{1-[(h/A)-k]^2\}^{\frac{1}{2}}, \tag{5}$$

$$u=(h/v\gamma)\pm(h^2-\gamma)^{\frac{1}{2}}/\gamma. \tag{6}$$

Wherein we have put $\gamma=(1/v^2)-1$. We note that $\gamma>0$ since we shall consider only the cases in which $v<1$. (See Fig. 1 for ranges of values of u and h.) The positive sign of the square root in Eq. (5) is to be taken when $\cos y>0$, negative otherwise.

There is ambiguity in Eq. (6) as to the choice of sign. Going back to Eq. (2) and making use of Eq. (6) and the expression for dx/dt, we find

$$dy/dx=(1/v)-dt/dx=\pm(h^2-\gamma)^{\frac{1}{2}}/+(u^2-1)^{\frac{1}{2}}.$$

It is now seen that the square root in Eq. (6) is to be taken positive when $dx/dt>v$, and negative otherwise. The change of sign thus takes place when $dx/dt=v$ and therefore when also $h^2-\gamma=0$.

J. C. Slater[2] has employed a different method for obtaining the relationship of momentum and phase (easily convertible to energy and phase), which is not adaptable for obtaining the relationship of phase and distance. His results are embodied in Eqs. (3) and (6) above.

We now shall integrate our equations once more to obtain the energy and phase as a function of distance. In order to do this, we go back to Eq. (2) replacing y by its value in terms of h as given in Eq. (3) and u by its value in terms of h as given by Eq. (6). In this way, it is found that

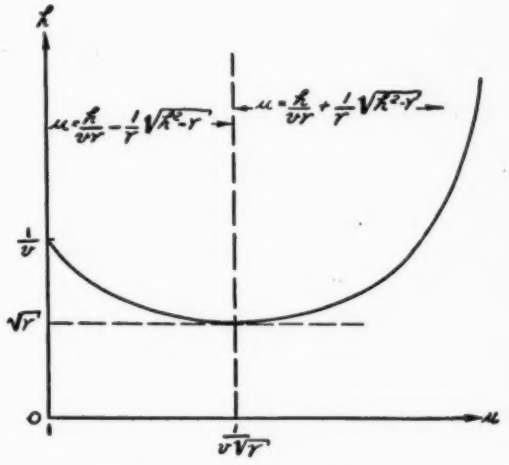


FIG. 1. Typical curve of h vs u.

$$\frac{(1/\gamma)h\,dh}{\pm\{[h^2-\gamma][A^2-(h-Ak)^2]\}^{\frac{1}{2}}}+\frac{dh}{Av\gamma\cos y}=dx. \tag{7}$$

The second term on the left side of this equation is seen from differentiating Eq. (3), to be equal to $dy/v\gamma$ and is, therefore, easily integrated to $(y/v\gamma)+$constant. The first term on the left side of Eq. (5) is reducible to elliptic integrals of the first and third kinds as well as an inverse trigonometric integral type. To get it into the standard forms, the standard substitution $h=(p+qz)/(1+z)$ is employed.[3] (We follow Edwards[3] notation.) By separating $(p+qz)/(1+z)$, appearing in the numerator, into partial fractions $q+(q-p)z/(1-z^2)+(p-q)/(1-z^2)$, we may now integrate Eq. (7) to

$$x-x_0=\frac{y}{\gamma v}\pm\frac{(q-p)}{\gamma(-HH')^{\frac{1}{2}}}\left\{\frac{q}{(-g')^{\frac{1}{2}}}\mathbf{F}\left[\sin^{-1}\left(\frac{z}{\sqrt{-g}}\right);\sin^{-1}\left(\frac{g}{g'}\right)^{\frac{1}{2}}\right]+\frac{(p-q)}{(-g')^{\frac{1}{2}}}\mathbf{\Pi}\left[\sin^{-1}\left(\frac{z}{\sqrt{-g}}\right);-g;\sin^{-1}\left(\frac{g}{g'}\right)^{\frac{1}{2}}\right]\right.$$
$$\left.+\frac{(p-q)}{[(1+g)(-g'-1)]^{\frac{1}{2}}}\sin^{-1}\left[\frac{(-g'-1)(-g-z^2)}{(-g'+g)(1-z^2)}\right]^{\frac{1}{2}}\right\} \tag{8}$$

(see Appendix for details), where $\mathbf{\Pi}$ indicates an elliptic integral of the third kind and $\mathbf{F}$ indicates an elliptic integral of the first kind, while x_0 is an integration constant. See Edwards[3] for details. This constant, x_0, must be re-evaluated whenever the plus or minus sign in Eq. (8) changes, in order to keep the distance traveled along the wave guide a continuous function of phase. The parameters p, q, H, H', g, g' are all constants determined once for all, for a given particle in a given section of the accelerator, by its initial energy and phase and the speed of the wave.

For the case of the separatrix, the condition is $-A+Ak=+\sqrt{\gamma}$, so that integrating,[4] we find

$$x-x_0=\frac{y}{v\gamma}\mp\frac{1}{(\gamma A\sqrt{\gamma})^{\frac{1}{2}}}\sinh^{-1}\left\{\frac{\sqrt{\gamma}(2A-h+\sqrt{\gamma})}{(A+\sqrt{\gamma})(h-\sqrt{\gamma})}\right\}^{\frac{1}{2}}$$
$$\pm\frac{1}{\gamma}\sin^{-1}\left(\frac{h-A}{A+\sqrt{\gamma}}\right). \tag{9}$$

If the field strength is large enough, the injected electron may be slowed to the point where $dx/dt=0$. When and if this should happen, $u=1$, and thus $h=1/v$

[2] J. C. Slater, Revs. Mod. Phys. **20**, 505, 508 (1948).

[3] J. Edwards, *The Integral Calculus II* (The Macmillan Company, New York), pp. 582, 583.

[4] Greenhill, *Elliptic Functions* (The Macmillan Company, New York), p. 58, II, (ii).

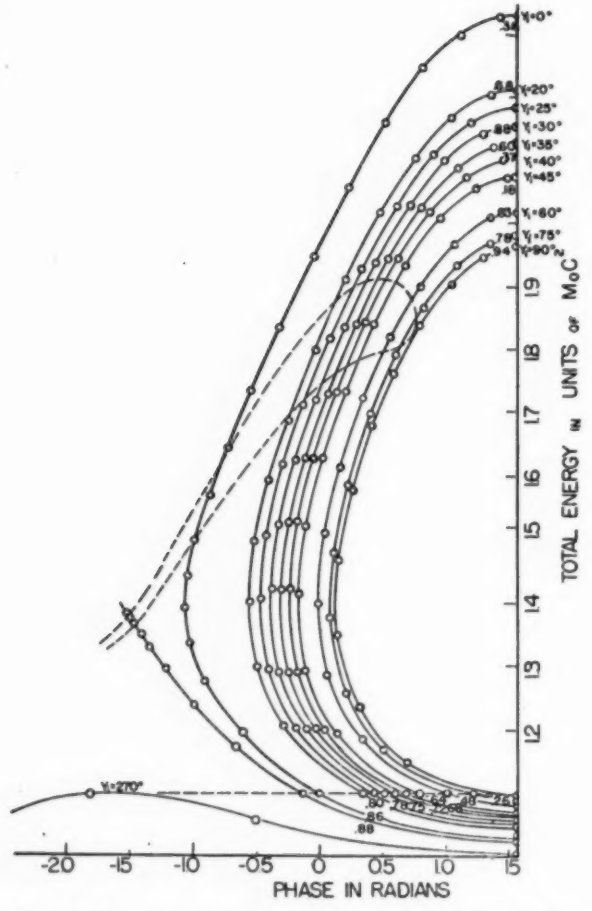


FIG. 2. Phase energy space, $v=0.70$. Distance corresponding to space between two successive circles on a given curve is 1.00 units of $\lambda/2\pi$, unless otherwise indicated. Note: y_i is the phase, in degrees, of injection into the first section of the Purdue Accelerator.

(as seen from the definition of h (Eq. (4)), and the definition of u). In order for this to be possible, $A+Ak$, the upper limit of h (that is a_1), must also be at least as large as $1/v$. Now, since $Ak=h_i-A\sin y_i$, where the subscript i refers to the injection values, the maximum value of Ak occurs for injection phase $y_i=-90°$ and this value of Ak is just h_i+A. Thus, in order for the speed of the electron to become zero; $A+Ak=h_i+2A$ (in the optimum case), and thus $h_i+2A\geqq 1/v$. Calculations for the Purdue Accelerator show that this does not occur for the particles considered, that is those particles injected with kinetic energy of 50 kev.

CONCLUSIONS

The graphs of Fig. 2 show the complete orbits that the electrons would have traversed if the sections of the Purdue Accelerator had not been cut off where indicated. All curves are labeled according to their initial injection phases, y_i, of the electron injected into the first section, for which $v=0.7$.

In particular, Figs. 2 and 3, both of which are completely symmetric about the vertical line, phase$=y=\pi/2$, show the energy-phase space for the cases considered, $v=0.7$, $A=0.10898$, and $v=0.96$, $A=0.082561$ (the first two sections of the Purdue Accelerator). The dotted curves show the exact places that the initially injected particles, all of 50 kev, find themselves in the energy phase space at the beginnings and the ends of both sections. Figure 8 shows the energy distribution of electrons at the end of the second section of the Purdue Accelerator. Examination shows that about 90/360 =one quarter of the injected electrons are caught by the accelerator and accelerated to a rather narrow energy range of $u=6.7$, or *kinetic* energy$=2.9$ Mev.

These two phase space diagrams, Figs. 2 and 3, may be used to see what happens if the electrons are not all injected into the first section with the same energy. In such a case, they would not be represented at the injection end in Fig. 2 by a horizontal line, but by some sort of area. This area would move up in the phase space in such a way that each point inside the area would move through a given amount of phase

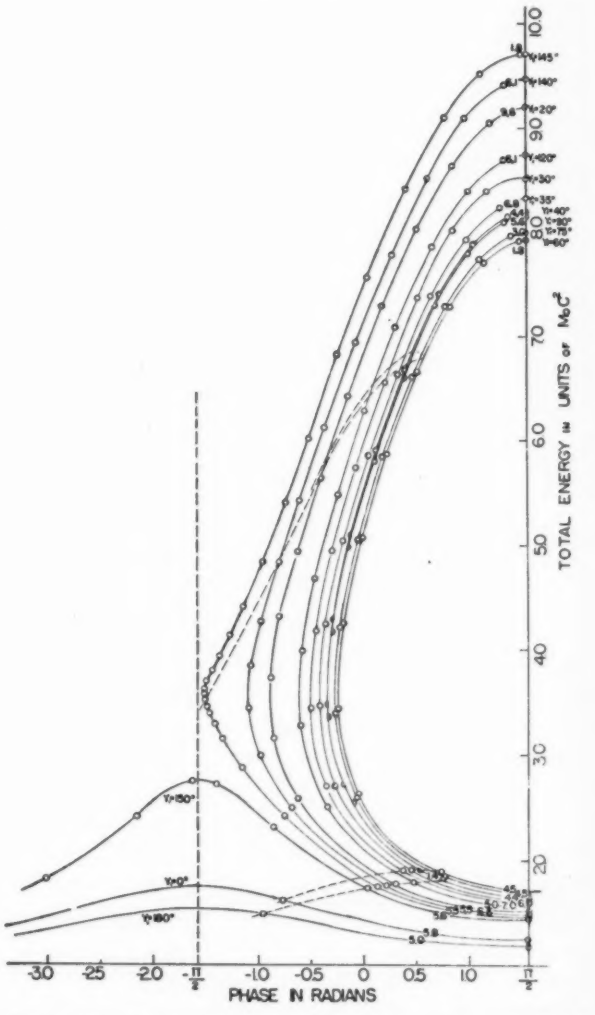


FIG. 3. Phase energy space, $v=0.96$. Distance corresponding to space between two successive circles on a given curve is 10.00 units of $\lambda/2\pi$, unless otherwise indicated (also see note to Fig. 2).

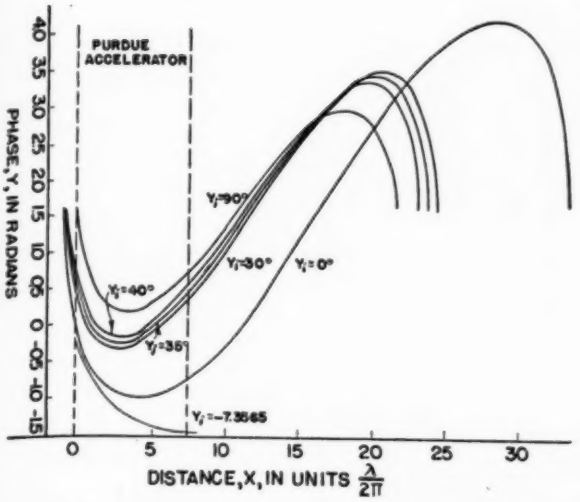


FIG. 4. Phase *vs* distance, $v=0.70$ (see note to Fig. 2).

space corresponding to the length of the wave-guide section in physical space, thus yielding a new area representing the phase-energy spread of the particles at the other end of the section. What comes out of the second section is obtained in a similar manner.

CONSTRUCTION OF GRAPHS

Putting into Eq. (8) the values appropriate to the Purdue accelerator, there can be tabulated the calculated values of x, the corresponding value of u, and the corresponding value of y. From these, Figs. 4–7 are plotted. The second half of these graphs have been obtained from the symmetry property of the phase and energy plots about $y=\pi/2$, and the energy maximum (which occurs at $y=\pi/2$), respectively.

The curves in Figs. 2 and 3 were drawn up from the corresponding sets of values of energy and phase. Then

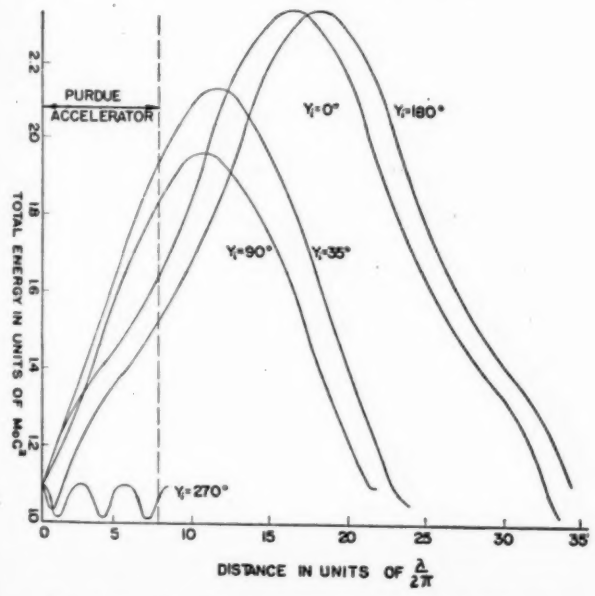


FIG. 5. Energy *vs* distance, $v=0.70$ (see note to Fig. 2).

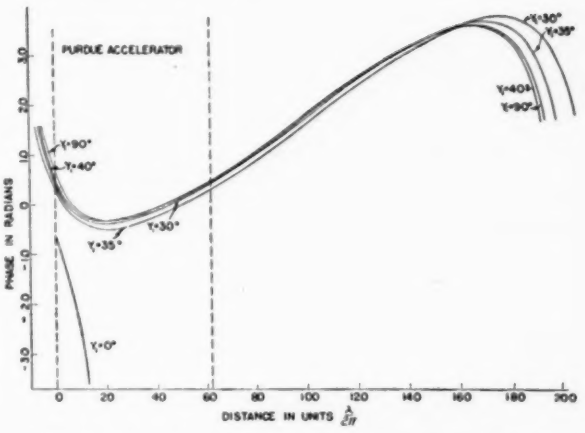


FIG. 6. Phase *vs* distance, $v=0.96$ (see note to Fig. 2).

the points corresponding to certain increments of the distance, x, were circled for the case of each particle by means of graphs similar to Figs. 4 or 5 and Figs. 6 or 7; the choice depending upon where the greatest accuracy is obtainable.

Figure 8, the final total energy of the various injected particles *vs* initial injection phase into the *first* section (all of 50 kev kinetic energy) was obtained from the values of u *vs* x for each particle directly, by simply looking up the value of the total energy of the particle at the end of the second section, for which $x=61.826$ units $=104.34$ cms.

ACKNOWLEDGMENT

The authors wish to thank Dr. F. F. Riecke for making possible the use of the funds of the Linear Accelerator Project of Purdue University, under the contract of the Office of Naval Research, for the calculations of this problem.

APPENDIX

In Eq. (7), after the substitutions of Edwards[3] have been performed, there will occur an integral of the form

$$\int zdz(1-z^2)^{-1}(z^2+g)^{-\frac{1}{2}}(z^2+g')^{-\frac{1}{2}}.$$

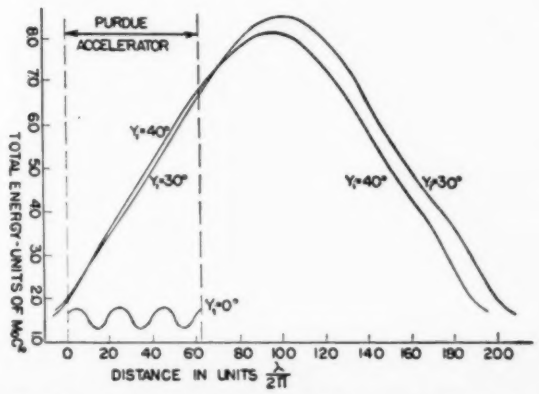


FIG. 7. Energy *vs* distance, $v=0.96$ (see note to Fig. 2).

To get this into standard form, introduce $s=z^2$, $ds=2zdz$, and remembering that $0<z^2<-g\sqrt{1}\sqrt{-g'}$, and one finds

$$(1/2)\int ds(1-s)^{-1}(-g-s)^{-\frac{1}{2}}(-g'-s)^{-\frac{1}{2}}.$$

This is integrated[5] to

$$\frac{-1}{[(-g'-1)(1+g)]^{\frac{1}{2}}}\sin^{-1}\left\{\frac{(-g'-1)(-g-z^2)}{(-g'+g)(1-z^2)}\right\}^{\frac{1}{2}}.$$

In Eq. (7) we will have also an integral of the form

$$\int dz[(z^2+z)(z^2+g')]^{-\frac{1}{2}}$$

$$=\int dz(gg')^{-\frac{1}{2}}[(1+z^2/g)(1+z^2/g')]^{-\frac{1}{2}}.$$

Now set $t=z/\sqrt{-g}$, then $t^2=z^2/-g\leqq 1$. So, set $t=z/\sqrt{-g}=\sin\phi$, and we find

$$(-g')^{-\frac{1}{2}}\int d\phi[1-(g/g')\sin^2\phi]^{-\frac{1}{2}}$$

$$=(-g')^{-\frac{1}{2}}\mathbf{F}[\sin^{-1}(z/\sqrt{-g});\sin^{-1}(g/g')^{\frac{1}{2}}],$$

where $\mathbf{F}$ denotes an elliptic integral of the first kind.

In Eq. (7), there will be another integral to be evaluated. It is of the form

$$\int dz(1-z^2)^{-1}[(z^2+g)(z^2+g')]^{-\frac{1}{2}}.$$

Introducing again $\sin\phi=z/\sqrt{-g}$, one obtains

$$(-g')^{-\frac{1}{2}}\int d\phi(1+g\sin^2\phi)^{-1}[1-(g/g')\sin^2\phi]^{-\frac{1}{2}}.$$

This is now in the standard form for the elliptic integral of the third kind. Specifically, it is equal to (in Legendre's notation)

$$(-g')^{-\frac{1}{2}}\mathbf{\Pi}[\sin^{-1}(z/\sqrt{-g})=\phi;\quad g=n;\quad (g/g')=\kappa^2].$$

The complete integral of the third kind, that is for which $\phi=\pi/2$, is to be found tabulated in tables.[6] There are no tables, at the present writing, for the incomplete integral of the circular case.[7,8] That this case is the circular type is due to the fact that

$$\alpha=(1+n)/[1+(\kappa^2/n)]=g'(1+g)/(1+g')>0,$$

since $-g'>1>-g>0$.

For the numerical calculation of this incomplete integral, binomial series expansions of $(1-z^2)^{-1}$ and $(z^2+g')^{-\frac{1}{2}}$ are to be used and integrated term by term, noting that $z^2<-g<1<-g'$.

For orbits near the separatrix, $-g$ is nearly unity, and at those points where $z/\sqrt{-g}$, is or is near unity, the convergence of these expansions is very slow. This occurs for example, in the neighborhood of the self-crossing point of the separatrix where the phase is about $-90°$. In order to obtain points in this region, the addition theorem[8] for elliptic integrals of the third kind is used

$$\mathbf{\Pi}(\phi)+\mathbf{\Pi}(\psi)=\mathbf{\Pi}(\mu)$$

$$+\frac{1}{\sqrt{\alpha}}\tan^{-1}\left[\frac{g(\alpha)^{\frac{1}{2}}\sin\phi\sin\psi\sin\mu}{1+g-g\cos\phi\cos\psi\cos\mu}\right],\quad \text{(A1)}$$

in which

$$\mathbf{F}(\phi)+\mathbf{F}(\psi)=\mathbf{F}(\mu),\quad \text{and}\quad \alpha=(1+g)g'/(1+g').\quad \text{(A2)}$$

By setting $\mu=\pi/2$ and ψ to some conveniently small angle, ϕ comes out of Eq. (A2) close to, but less than $\pi/2$. $\mathbf{\Pi}(\psi)$ is now calculated by means of the rapidly converging series above (since ψ is small) and $\mathbf{\Pi}(\mu)$ is the complete integral of the third kind, obtainable from tables. Thus, $\mathbf{\Pi}(\phi)$ is obtained (by means of Eq. (A1)) more easily for values of the argument near $\pi/2$ where the series expansion converges too slowly.

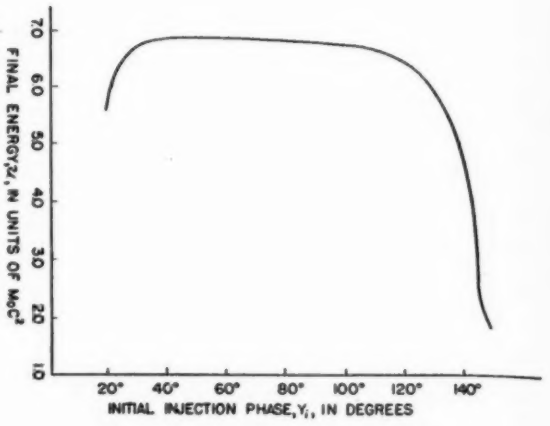


FIG. 8. Final energy *vs* initial phase.

[5] Greenhill, *Elliptic Functions*, p. 58, No. II, (iv).
[6] C. Neuman, J. Math. Phys. II, 2, p. 127*ff*.
[7] A. Greenhill, see reference 5, pp. 193–194.
[8] *Mathematical Tables and Aids to Computation* (October, 1948), Vol. III, 24, p. 243.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

The Electrical Resistance of Binary Metallic Mixtures

ROLF LANDAUER
*Lewis Laboratory, National Advisory Committee For Aeronautics, Cleveland, Ohio**
(Received March 10, 1952)

The existing theories of the resistivity of mixtures assume regular arrangements of the two components, rather than random mixtures. A theory for a random mixture is given, based on the assumption that each crystal acts as if surrounded by a homogeneous medium whose properties are those of the mixture. Comparisons with experiment are made. The experimental data that have been examined fall roughly into two classes. One class consists of mixtures, where the variation of resistivity with composition disagrees violently with this theory, making it clear that the assumptions made are completely inapplicable. The remaining class consists of mixtures which generally agree well with the theory.

INTRODUCTION

THE problem of the resistance of binary metallic mixtures has attracted attention for almost one hundred years. Despite this, only a few theories have been suggested, and these have been tested by only a few comparisons with experiment. Some of these comparisons turn out to be insignificant upon close examination.

We have in mind a situation where small pieces of a *homogeneous* phase, denoted by 1, are mixed with similar pieces of a second phase denoted by 2. It is supposed that the two types of crystals completely fill space and furthermore, that there is good contact between the crystals so that the potential drop across the grain boundaries may be neglected. Furthermore, we are considering the typical mixture in which the crystals can be seen microscopically and are therefore large enough so that there is no question about the applicability of a macroscopic conductivity theory.

EXISTING THEORIES

Different mixtures show a variety of shapes and arrangements for the components. No single theory can be expected to cover all cases. We shall review here briefly the most important models that have been used in arriving at various theories.

Let x_1 stand for the fraction of the total volume occupied by material 1, let x_2 stand for the fraction of the total volume occupied by material 2, let σ_1 and σ_2 be the respective conductivities, and let ρ_1 and ρ_2 be the resistivities. An obvious guess, frequently used in metallurgical literature, is

$$\rho_m = x_1\rho_1 + x_2\rho_2, \tag{1}$$

where ρ_m is the resistivity of the mixture. This would be correct if the materials were arranged in alternate flat layers, perpendicular to the direction of current flow. In such a model the current cannot avoid the regions of high resistance, as it will do if the regions are mixed at random. Therefore, Eq. (1) gives an upper limit for the resistivity. The alternate guess,

$$\sigma_m = 1/\rho_m = x_1\sigma_1 + x_2\sigma_2, \tag{2}$$

gives an upper limit for the conductivity. Equation (2) corresponds to a model in which the layers are parallel to the direction of current flow. The current can therefore flow straight through the regions of low conductivity preferentially, without having to follow a warped course, as will be the case in a random mixture.

A guess, which agrees better with the measured values of the resistance of metallic mixtures, was put forth by Lichtenecker[1]:

$$\rho_m = \rho_1^{x_1}\rho_2^{x_2}. \tag{3}$$

In comparing the results of the various existing theories and guesses with experimental results it becomes apparent that any reasonable method of interpolating between σ_1 and σ_2, or between ρ_1 and ρ_2 gives good results if the two conductivities are of the same order of magnitude. A serious test of such theories can only be obtained by considering the case where the conductivities σ_1 and σ_2 differ very appreciably.

It can easily be seen that Eq. (3) does not meet this test. If there is only a very small amount of material 2 with $\rho_2=0$, Eq. (3) gives $\rho_m=0$. Actually, a few small particles of perfect conductor would not alter the conductivity of a medium appreciably, since most of the current's path would still be in the medium of finite conductivity. Similarly, a small trace of 2 with $\rho_2=\infty$ would give $\rho_m=\infty$, which is again unreasonable.

Lord Rayleigh[2] has considered the problem in a paper whose primary purpose was to test the Lorenz-Lorentz theory. Rayleigh assumes that material 2 exists in the form of spheres or cylinders, forming a regular rectangular array which is embedded in material 1. In this model there exist lines of current flow which stay entirely within material 1 and are never interrupted by meeting an obstacle made of material 2. On the other hand, current can never flow from a piece of material 2 to another piece of the same material without traversing a region of material 1. Therefore, even if there are equal volumes of 1 and 2, it will be primarily the conductivity of material 1 that matters. Actually, we would expect that in most such mixtures the two components should be of equal importance. Rayleigh's

* Now with the International Business Machines Corporation, Poughkeepsie, New York.

[1] K. Lichtenecker, Physik. Z. **25**, 225 (1924).
[2] Lord Rayleigh, Phil. Mag. **34**, 481 (1892).

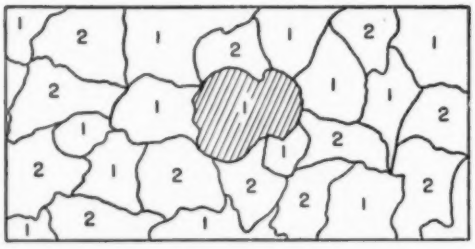


FIG. 1. The shaded crystal of type 1 is surrounded by crystals of both types which are imagined to be replaced by a single medium of uniform conductivity.

type of treatment is therefore accurate only if the amount of material 2 is small compared to that of 1 or if the situation is one in which material 1 actually always envelopes material 2.

Since Rayleigh's method can be used when there is a large percentage of one phase and a small percentage of the second phase, it can be used at both ends of the range of mixture in a given binary system. If we had a simple way of interpolating between the two ends, we would then have a reasonable approximation for the whole range. It will turn out that the equation to be developed in this paper will do exactly this. (This property will, however, not be the main justification given for the equation.)

A series of theoretical papers by Lichtenecker[3–6] using Matthiessen's[7] measurements give extensive calculations based on complicated Ohm's law approximations. A structure in which one material is regularly spaced and imbedded in another material is always assumed. Lichtenecker considers the systems Pb-Cd, Pb-Sn, Sn-Cd, Sn-Zn, and Cd-Sn. Of these five systems only the comparisons with the experimental measurements for Pb-Cd can be considered significant. In this case Lichtenecker obtains good agreement. Sn-Cd, Sn-Zn, and Cd-Sn will be discussed in a later section. Lichtenecker does get good agreement with Matthiessen's measurements on Pb-Sn. This agreement is, however, fortuitous. Matthiessen's measurements for different compositions in the same system were taken at different temperatures. To obtain a meaningful comparison with the theory, the measurements must be corrected to give the resistance as a function of composition at a fixed temperature. This correction was not applied by Lichtenecker,[8] at least in the case of Pb-Sn, where the range of temperatures involved is larger than in the other systems. If the correction is applied, the experimental curve becomes more irregular, and the agreement with any reasonable theory, including the one to be presented here, is poor.

Calculations similar to those of Lichtenecker were done by Guy S. son Frey.[9] son Frey stresses the importance of the model and considers three situations. In one case the better conductor is assumed to form a continuous network in which the poorer conductor is imbedded. In another case the poorer conductor is continuous, and the better conductor is imbedded in it. In the remaining case both materials are assumed continuous. son Frey considers the systems Pb-Cd, Pb-Sn, Sn-Cd, and Sn-Zn. The remarks that apply to Lichtenecker's comparisons for these systems apply equally to son Frey's.

son Frey also discusses Bi-Sn and Ag-Bi. For Bi-Sn he assumes that both materials form an unbroken connected network. The agreement he obtains with the experimental results makes this assumption plausible. An alternate possibility will, however, be given later in this paper (see Fig. 5).

In the Ag-Bi system the conductivity drops very rapidly with the addition of only a small amount of bismuth to the silver. This is usually explained[10] by assuming that up to about 5 percent bismuth (by weight) is soluble in silver. If this usual explanation is correct, the resistivity curve should have a bend at the composition of silver saturated with bismuth. son Frey instead explains the drop in conductivity by assuming that the small amount of poorly conducting bismuth forms a shell surrounding the silver crystals and in this way obtains a theoretical curve without a sharp bend. The experimental data do not contain enough measurements to really show whether there is a sharp bend or not. It should, however, be noted that the experimental value of conductivity, which Matthiessen obtained for a mixture with 16 percent bismuth by weight, is entered incorrectly on son Frey's Fig. 12. If this error is corrected, the experimental measurements make the existence of a bend in the resistivity curve much more likely, and the agreement with son Frey's theory becomes poorer.

A comparison of Rayleigh's theory with Matthiessen's[7] measurements on Pb-Cd is given by Pirani and Runge.[11]

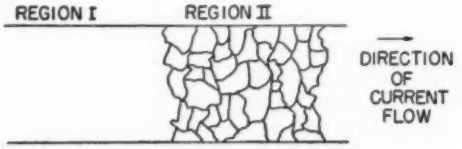


FIG. 2. Region II is the mixture under consideration. Region I is a homogeneous medium with the same conductivity as the mixture.

[3] K. Lichtenecker, Physik. Z. **10**, 1005 (1909).
[4] K. Lichtenecker, Physik. Z. **25**, 169 (1924).
[5] K. Lichtenecker, Physik. Z. **25**, 193 (1924).
[6] K. Lichtenecker, Physik. Z. **25**, 666 (1924).
[7] A. Matthiessen, Pogg. Ann. **110**, 190 (1860).
[8] Lichtenecker (see reference 3) claims that he has corrected the measurements to 20°, but this is not in agreement with the values shown in his Fig. 2 of reference 3. In reference 1 diagram IV, Lichtenecker also omits two of Matthiessen's experimental points. One is in the Pb-Sn series, the other is in the Zn-Sn series.

[9] Guy S. son Frey, Z. Elektrochem. **38**, 260 (1932).
[10] All the information on phase diagrams relevant to this paper can be found in M. Hansen, *Aufbau der Zweistofflegierungen* (Julius Springer, Berlin, 1936).
[11] M. Pirani and J. Runge, Z. Metallkunde **16**, 183 (1924).

A very general theoretical discussion of the problem has also been given by Wiener,[12] and a review paper was written by Lichtenecker.[13]

THEORY

We shall derive an equation for σ_m which will involve the two components in a symmetrical fashion. Consider Fig. 1, which shows the shaded region of material 1, surrounded by an assortment of other regions of both types.

Our primary assumption will be that we can treat the neighborhood of the shaded region as if it were a uniform medium, which has the conductivity σ_m that characterizes the mixture. This will only be true if there is no correlation between the positions of the two types of regions. (If a region of type 1 is preferentially surrounded by either other regions of type 1 or by regions of type 2, we cannot assume the shaded region to be imbedded in a uniform medium.) The conductivity in the shaded region itself is σ_1. We shall therefore treat the shaded region as a single inclusion of conductivity σ_1 in an infinite uniform medium of conductivity σ_m. Let $\mathbf{E}_m$ be the average electric field in the mixture and therefore also the field that exists far away from the shaded region. Let us also assume that the shaded particle is spherical and has radius a. Then there will be a charge deposited at the surface of the shaded region, which according to elementary electrostatics has a dipole moment

$$\mathbf{E}a^3(\sigma_1-\sigma_m)/(\sigma_1+2\sigma_m). \tag{4}$$

If there are N_1 such regions per unit volume, then the polarization will be

$$\mathbf{P}_1=N_1a^3\mathbf{E}(\sigma_1-\sigma_m)/(\sigma_1+2\sigma_m) =x_1\mathbf{E}(\sigma_1-\sigma_m)/(\sigma_1+2\sigma_m). \tag{5}$$

Similarly, the regions of type 2 will produce a polarization

$$\mathbf{P}_2=x_2\mathbf{E}(\sigma_2-\sigma_m)/(\sigma_2+2\sigma_m). \tag{6}$$

Now consider the situation shown in Fig. 2. Region *II* is the mixture under consideration with conductivity σ_m. Region *I* is assumed to be a completely homogeneous conductor with the same conductivity σ_m. Let $\mathbf{E}_I$ be the field in *I*, and $\mathbf{E}_{II}$ the field in *II*. Since regions *I* and II carry the same current and have the same conductivity,

$$\mathbf{E}_I=\mathbf{E}_{II}. \tag{7}$$

The electric field, however, obeys

$$\text{div}\mathbf{E}=-4\pi\,\text{div}\mathbf{P}. \tag{8}$$

Hence,

$$\mathbf{E}_{II}-\mathbf{E}_I=-4\pi(\mathbf{P}_1+\mathbf{P}_2), \tag{9}$$

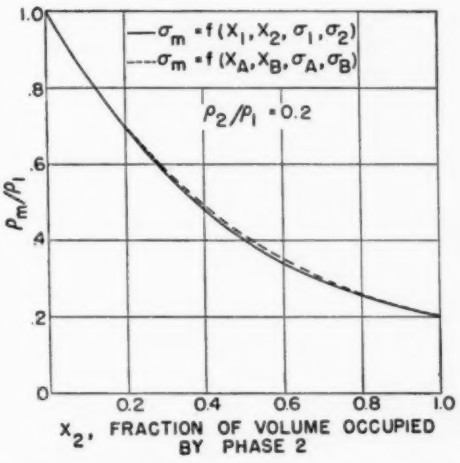


FIG. 3. The solid curve is computed on the assumption that the pure phases 1 and 2 are mixed. The dashed curve assumes that the material containing 0.20 (by volume) of material 2 retains its character and is not broken up into its original components when it is mixed with additional amounts of material 2 to form mixtures which have more than 20 percent of 2 by volume.

and therefore combining (7) and (9),

$$\mathbf{P}_1+\mathbf{P}_2=0. \tag{10}$$

This gives the equation for σ_m:

$$x_1(\sigma_1-\sigma_m)/(\sigma_1+2\sigma_m)+x_2(\sigma_2-\sigma_m)/(\sigma_2+2\sigma_m)=0, \tag{11}$$

whose solution is

$$\sigma_m=\tfrac{1}{4}\{(3x_2-1)\sigma_2+(3x_1-1)\sigma_1 +[((3x_2-1)\sigma_2+(3x_1-1)\sigma_1)^2+8\sigma_1\sigma_2]^{\frac{1}{2}}\}. \tag{12}$$

This derivation has assumed that all the regions are spherical and therefore does not directly apply to a mixture which has needle or disk shaped inclusions. Equation (12) can still give a rough estimate of the conductivity, since the direction of maximum polarizability of the disks or needles will, in general, occur equally likely at all orientation with respect to the field. An exact treatment of the cases of needles (or disks) would be difficult, since the orientation of nearby needles (or disks) will usually tend to be correlated. Even more complicated would be the treatment of the dendritic structures frequently found in alloys.

METHOD OF APPLICATION

In some of the cases to be discussed, the conductivities of the terminal members of the series, phases 1 and 2, are not known. In others the conductivities of the pure phases or of the pure phases and the phases that are close to pure do not fit into the experimental curve, which otherwise is smooth and regular. In these cases we cannot expect to predict σ_m from σ_1 and σ_2. In such cases only the range of mixtures in which the experimental measurements fit on a reasonable curve have been considered. Let *A* denote the mixture in this range, which is richest in phase 1 and for which the resistivity is known. Let *B* denote the mixture in

[12] O. Wiener, Abhandlungen der Mathematish-Physischen Klasse der Koenigl. Saechsischen Gesellschaft der Wissenschaften 32, 507 (1912).

[13] K. Lichtenecker, Physik. Z. 27, 115 (1926).

this range which is richest in phase 2 and for which the resistivity is known. Now physically, all mixtures intermediate between A and B can be prepared from A and B. We shall use this as a basis for our calculations. Instead of regarding σ_m as a function of x_1, x_2, σ_1, and σ_2, we shall regard it as a function of x_A, x_B, σ_A, and σ_B. Here x_A and x_B represent, respectively, the volumes of A and B which have to be mixed to obtain the intermediate mixtures. σ_A and σ_B are the conductivities of A and B, respectively.

Our modified theory is not strictly correct. If the intermediate mixtures in the range in which we are interested really consisted of large patches of crystals in which the ratio of 1 to 2 was as in A and other large patches in which the ratio of 1 to 2 was as in B, then our procedure giving $\sigma_m=f(x_A, x_B, \sigma_A, \sigma_B)$ would be correct, for then we could consider the large patches as the basic components. Actually, however, in preparing the intermediate mixtures from the melt of A and B, it is the individual crystals of 1 and 2 that will be mixed, and the final mixture has nothing about it that is particularly characteristic of A and B. The procedure giving

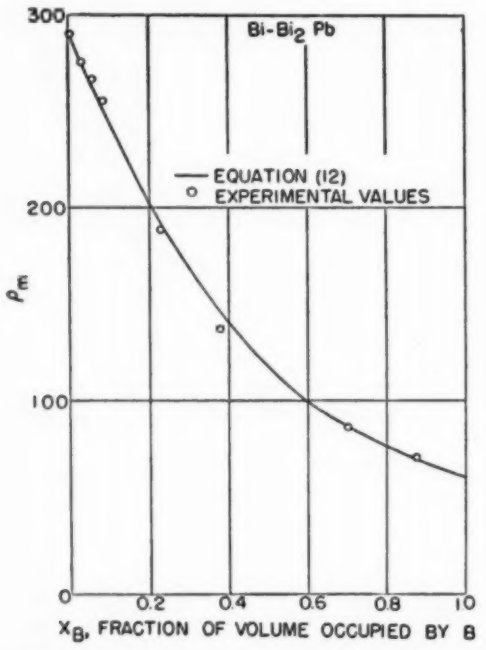


FIG. 4. The experimental values were measured by Herold [W. Herold, Z. anorg. u. allgem. Chem. **112**, 131 (1920)] at 0°C. A small amount of lead is apparently soluble in bismuth and affects its resistivity extremely critically. A is therefore taken to contain 4 percent lead by weight. The solubility of Bi in Bi_2Pb is not known but is assumed very small, so that the composition of this saturated phase can be taken to be that of pure Bi_2Pb. The resistivity of this saturated phase, which is taken to be B, is not known. An assumed value of 60 μohm cm gives good agreement with the experimental data and has been used here. Herold measured the resistivities both before and after annealing. The values measured before annealing fit more smoothly on a simple curve and have been used here. Matthiessen's (see reference 7) measurements on Bi-Pb disagree violently with Herold's and cannot possibly be explained by any mixture theory. In Matthiessen's measurements the resistivity at the bismuth end drops with the addition of lead at a rate faster than permitted by Eq. (2).

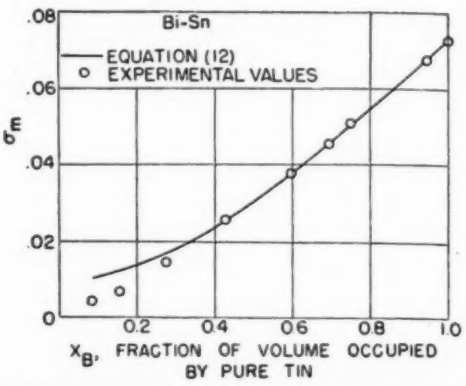


FIG. 5. The conductivity in (μohm cm)$^{-1}$ is shown here to emphasize the tin end of the curve. The experimental points represent Matthiessen's (see reference 7) measurements corrected to 26°C. At the extreme bismuth end of the curve a number of experimental points are omitted. B is taken to be pure tin. A is taken to be bismuth saturated with tin. The composition of A is taken to be negligibly different from pure bismuth. The resistivity of A is assumed to be 120 μohm cm. (Other assumptions, for instance, that 5 percent of tin, by weight, is soluble in bismuth and that the resistivity of this saturated solution is 96 μohm cm, give equally good agreement.) The poor agreement at the bismuth end may come from the fact that the experimentally prepared mixtures near the bismuth end do not represent equilibrium states. The violent disagreement at the bismuth end between differential experimental measurements on this system support this viewpoint [see reference 15 and F. A. Schulze, Ann. Physik **9**, 555 (1902); C. L. Weber, Ann. Physik **34**, 576 (1888)]. son Frey (see reference 9) obtains better agreement at the bismuth end, by assuming that both components of the mixture form a continuous network. His agreement with the experimental values near the tin end is, however, poorer than ours.

$\sigma_m=f(x_A, x_B, \sigma_A, \sigma_B)$ is nevertheless a very good approximation. This can be seen from Fig. 3, which describes a case in which $\rho_2/\rho_1=0.2$. The lower curve shows ρ_m/ρ_1 computed according to Eq. (12) from x_1, x_2, σ_1, and σ_2. The upper curve is computed according to Eq. (12) from x_A, x_B, σ_A, σ_B. A is taken to be a mixture containing 20 percent by volume of 2. B is taken to be the pure phase 2. If instead, B is also taken to be a mixture, the two curves approach each other rapidly and agreement becomes even better.

EXPERIMENTAL COMPARISONS

In many phase diagrams the regions in which mixtures occur are small, and there are not enough measurements in the region of mixture to permit comparison with theory. In other cases the measurements are so obviously incompatible with any reasonable theory of mixtures, that there is not much purpose in a comparison. Only cases for which there is some hope of an explanation along present lines will be discussed below.

Figures 4–10 give comparisons for the systems Bi-Bi_2Pb, Bi-Sn, Cd-Pb, Cu-Fe, Cu_2Sb-Sb, Mg_2Pb-Pb, Pb-Sb. These were the only ones, among all the measurements considered, where a theory of random mixtures can possibly explain the measurements. Among these seven cases it can be seen that five give excellent agreement. Only the system Mg_2Pb-Pb really fails to fit the theory. Bi-Sn fits the theory except near the bismuth

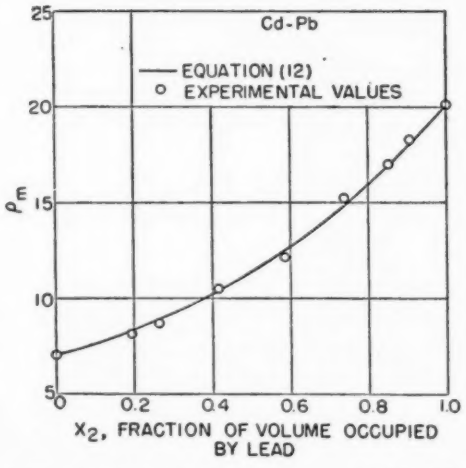


FIG. 6. The experimental values were measured by Matthiessen (see reference 7) and have been corrected to 21°C. It is assumed that pure cadmium and pure lead are the components of the mixture. Equally good agreement is obtained with the measurements of Beckman, as listed in the Critical Tables. [International Critical Tables (McGraw-Hill Book Company, Inc., New York, 1929), Vol. **VI**, pp. 156–206. Many of the experimental results mentioned in this paper can be found in the Critical Tables.] A slightly generalized form of Eq. (3) containing one adjustable constant is discussed by Lichtenecker (see reference 1) and gives agreement as good as that obtained with Eq. (12), son Frey's (see reference 9) theory, however, gives poorer agreement.

end, where the experimental measurements themselves show poor reproducibility. The other systems that were considered in detail but that are not shown on graphs will be briefly discussed. First of all there are Matthiessen's[7] measurements on Sn-Zn, Cd-Sn, Cd-Zn, and Pb-Sn. Matthiessen's measurements on Sn-Zn agree better with Eq. (2) than with any other theory. His measurements on Cd-Sn are too inaccurate to be taken very seriously, but on the average the intermediate mixtures have even larger conductivities than the limit represented by Eq. (2). Matthiessen's measurements on

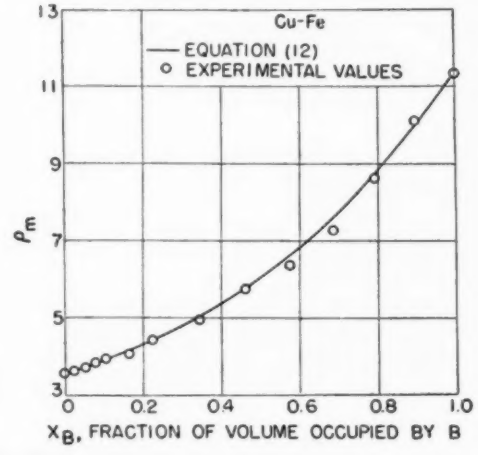


FIG. 7. The experimental values are the ones measured by Ruer and Fick [Ferrum **11**, 39 (1913)], as listed in the Critical Tables. *A* is taken to contain 2 percent by weight of Fe. *B* is taken to contain 10 percent Cu by weight. The experimental measurements at the extreme copper and iron ends do not fit smoothly into the remaining curve.

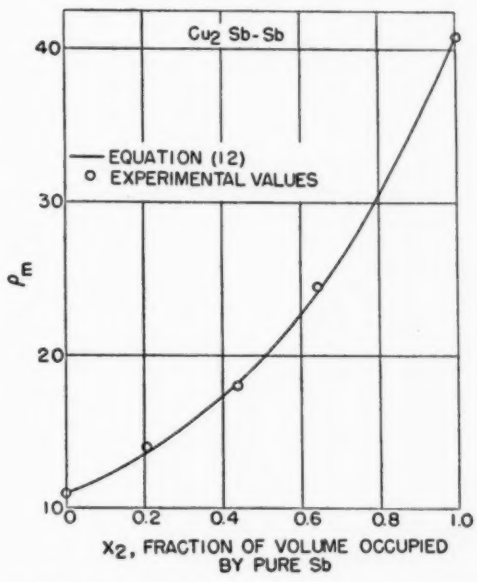


FIG. 8. The experimental values were measured by Stephens [E. Stephens, Phil. Mag. **7**, 161 (1929)] and have been corrected to 16°C. It is assumed that the pure phases Cu_2Sb and Sb have been mixed.

Cd-Zn and Pb-Sn are too inaccurate to make any comparisons significant. Roberts'[14] measurements on Pb-Sn are also too inaccurate, as are Bucher's[15] on Cd-Sn. Haken's[16] measurements on Sn-Te show that the resistivity of Sn rises much too fast with the addition of SnTe to agree with Eq. (12). Agreement with Eq. (1) is better, though not good. In the case of Ag-Bi, as measured by Matthiessen,[7] the resistivity at the silver end rises much too fast with the addition of bismuth, faster than the limit set by Eq. (1). Stepanow's[17] measurements on Mg-Sn show that if Mg_2Sn is first added to Sn, fair (but not good) agreement with Eq. (12) occurs. Near Mg_2Sn, however, the resistivity changes far too violently. Mg_2Sn is almost an insulator. It is clear that with the solution of only a small amount of Sn in Mg_2Sn the number of conducting levels (holes or electrons) in Mg_2Sn can be greatly multiplied. The usual phase diagram[10] indicates that very little Sn can dissolve in Mg_2Sn. One would therefore expect that over most of the range of mixtures Mg_2Sn-Sn, the mixtures would contain this saturated solution and therefore follow Eq. (12) with the simple modification that instead of using the conductivity of pure Mg_2Sn, the conductivity of Mg_2Sn saturated with Sn has to be used. Also one can expect an extremely sharp bend in the curve at the composition of the saturated solution, and an extremely rapid rise in resistance with the removal of the small amount of dissolved Sn. This expectation turns out to be incorrect. There is no sharp bend and the excessively rapid change in resistance

[14] Roberts, Phil. Mag. (ser. 5) **8**, 57 (1879).
[15] A. Bucher, Z. anorg. u. allgem. Chem. **98**, 97 (1916).
[16] W. Haken, Ann. Physik **32**, 291 (1910).
[17] N. J. Stepanow, Z. anorg. u. allgem. Chem. **78**, 1 (1912).

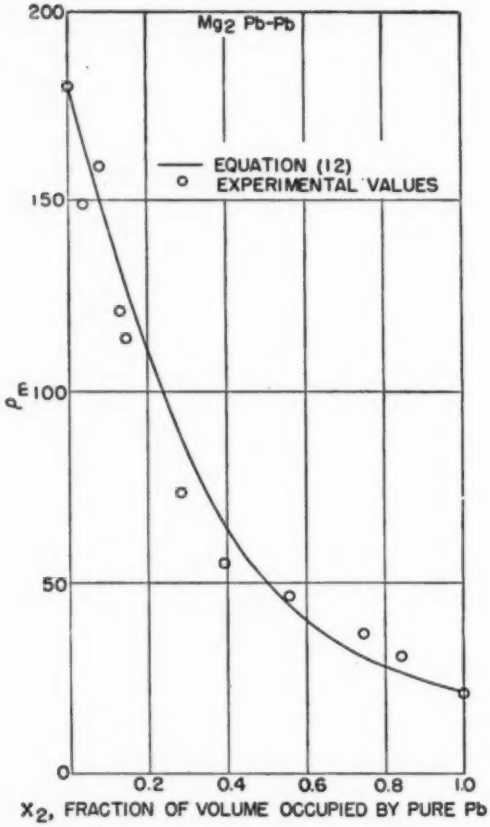


FIG. 9. The experimental values are from Stepanow's (see reference 17) measurements at 25°C. It is assumed that pure Mg_2Pb is mixed with pure Pb. The relatively sharp bend in the experimental curve near $x_2=0.4$ is puzzling.

lasts over a change in composition considerably greater than 1 or 2 percent. This is most plausibly explained by assuming that the measured samples do not represent equilibrium. Presumably, only when Mg_2Sn is mixed with a large proportion of Sn does the saturated solution have a chance to form. Similar situations probably prevail in some of the other Mg alloys, where intermetallic compounds form, as well as in some bismuth alloys. Lastly, the case of Stepanow's[17] measurements on $MgZn_2$-Mg may be mentioned. Despite their inaccuracy, it is clear that they agree much better with Eq. (1) than with Eq. (12).

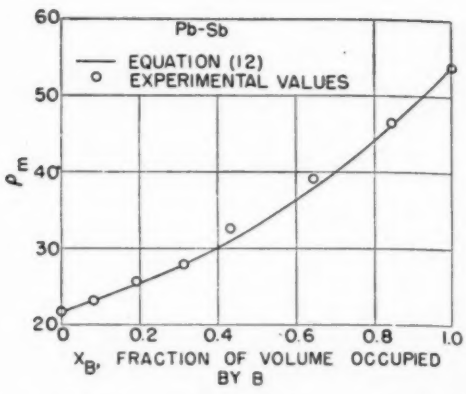


FIG. 10. The experimental measurements are those of Matthiessen (see reference 7) corrected to 24°C. The resistivity of pure lead is slightly out of line with the nearby experimental points. The resistivity of pure antimony quoted by Matthiessen is far too low (36.1 μohm cm at 18.7°C) to fit the curve. The theoretical curve has therefore been compuated by assuming that A consists of 2.31 percent Sb (by weight); B consists of 29.7 percent Pb (by weight). Except for pure Pb and Sb all other experimental points are shown.

EXPLANATION OF NUMERICAL DATA

Matthiessen's measurements were expressed by comparison to the conductivity of silver at 0°C. His results have been put on an absolute basis by assuming a resistivity of 1.54 μohm cm for the resistivity of silver at 0°C.

Temperature corrections were made, where necessary, by assuming a temperature coefficient of resistivity of 0.004/°C.

The density of intermetallic compounds, for example, Bi_2Pb, needed for the calculations was obtained from x-ray data.[18]

The units of resistivity, unless otherwise specified, are μohm cm.

ACKNOWLEDGMENTS

The author is indebted to Dr. George Kuczynski for suggesting this subject to him. He also wishes to thank Dr. S. L. Simon and Dr. P. Schwed for discussions.

[18] C. J. Smithells, Metals Reference Book (Butterworths, London, 1949), p. 180.

Grain Boundary Diffusion of Zinc in Copper

R. FLANAGAN AND R. SMOLUCHOWSKI
Carnegie Institute of Technology, Pittsburgh, Pennsylvania
(Received March 6, 1952)

In continuation of a previous study of grain boundary diffusion of silver in copper the grain boundary diffusion of zinc along grain boundaries of columnar copper has been investigated at various temperatures. The results confirm the influence of the angle between the grains on diffusion and indicate an angular dependence of the activation energy. The latter reaches very small or even negative values in a certain range of orientation in agreement with the proposed model.

THE question of the structure of grain boundaries and of grain boundary diffusion has been the subject of much interest. A study of diffusion of silver along grain boundaries of columnar copper[1] indicated that there is a pronounced variation of the structure of grain boundary with angle θ between the grains. At small angles the boundary is made up of individual dislocations and the diffusion is entirely controlled by volume diffusion. From a certain critical angle θ_c on, the grain boundary diffusion overtakes the volume diffusion and reaches a maximum at 45 degrees. It was concluded that rapid grain boundary diffusion occurs only in areas of highly distorted lattice. The purpose of the work here reported was to check whether the dependence of diffusion on angle between grains is found in other similar systems and to investigate the behavior of the activation energies with changing θ.

The copper-zinc system has been chosen since it is well known and it has a large solubility range which makes it possible to avoid complications due to precipitation.[2]

Experimental Procedure

The experimental procedure was similar to that used in I. The samples of columnar copper (of same purity and texture as in previous work) were enclosed in argon-flushed and evacuated Pyrex tubes with 30 percent Zn brass chips. This produced, upon heating, a constant zinc concentration on the outer surface of the columnar grains during the diffusion process. The diffusion runs were made at three temperatures: 550, 593, and 649°C; the corresponding times, 664, 332, and 146 hours, were chosen so as to give approximately the same penetration. The latter was estimated by guessing at a "reasonable" activation energy of 23,000 cal/mole.

After diffusion the samples were carefully measured, mounted in Bakelite, and consecutive layers approximately 0.025 mm thick were gradually removed by grinding and lapping. The zinc-rich zone surrounding the grain boundaries was made visible by etching in a mixture of equal parts of NH_4OH, H_2O_2, and H_2O. A study of the relative penetration along various grain boundaries can be made without actual knowledge of the composition at the etch limit. However, in order to get some idea of the concentration of zinc at the boundary of the observed zinc-rich zone a sample used by DaSilva and Mehl in a study[3] of volume diffusion in the Cu-Zn system was etched in the manner described above. It appeared that the etch limit lies between 10 and 15 percent Zn as estimated from a known penetration curve obtained from another similar sample. The error is due primarily to an uncertainty in the position of the penetration curve with respect to the etched sample.

The orientation of the columnar crystals was obtained by means of x-rays in the same manner as in I.

RESULTS

The results are shown in Fig. 1 in which the depth of penetration of zinc measured from the original interface is plotted as a function of the angle θ between the cube directions of the two grains. Depth of penetration is here defined as the average of the position of the last section

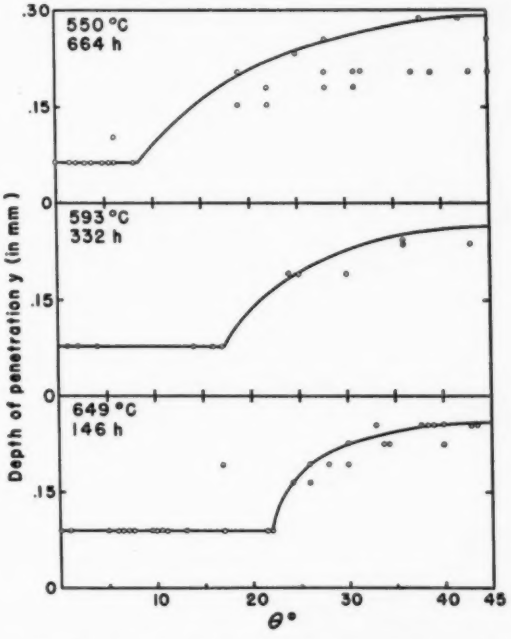


FIG. 1. Depth of penetration of zinc along grain boundaries of columnar copper as a function of orientation of the grains.

[1] M. R. Achter and R. Smoluchowski, J. Appl. Phys. **22**, 1260 (1951) [referred to as "I"]; Phys. Rev. **83**, 163 (1951).
[2] M. R. Achter and R. Smoluchowski, J. Appl. Phys. **23**, 373 (1952).
[3] L. DaSilva and R. F. Mehl, J. Inst. Metals **191**, 155 (1951).

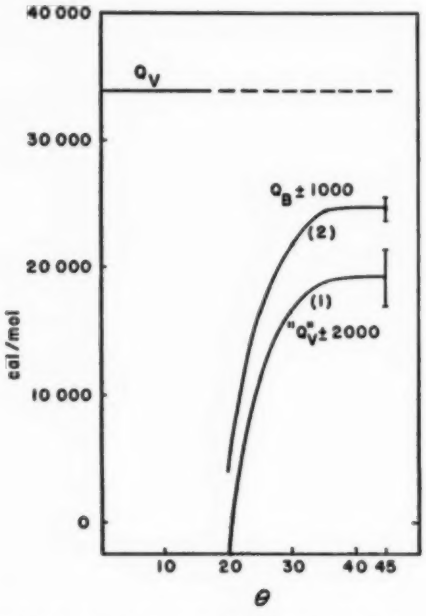


FIG. 2. Activation energies calculated from the volume diffusion Eq. (1) and from the Fisher solution (2).

where the zinc was still visible at a particular grain boundary and the next section.

As discussed in I much scatter of the experimental points is due to imperfect columnar structure. On the basis of Laue diagrams only those pairs of grains were considered which were columnar within 12 degrees. The results are of the same general character as those obtained in I: constant volume diffusion up to a certain critical angle and then a rapid rise of the grain boundary diffusion reaching a maximum at 45 degrees. In considering the apparent scatter of the points it should be kept in mind that the elimination of strongly noncolumnar grains does not remove all difficulties. The three additional factors are (a) deviation of the grain boundary from a position symmetrical with respect to the two grains, (b) change in the position of the plane containing the two almost parallel cubic directions with respect to the plane of the grain boundary and (c) the formation of cusps at certain particular angles.

The first factor presumably decreases the diffusion rate though this is not quite certain (see I). The second factor, which is rather difficult to measure accurately, affects the character and orientation of dislocations and may lower the rate of diffusion.[1] Finally the cusps are known to show up as abnormally low diffusion rates[1] because of a particularly good fit between the two grains.

Since any irregularity introduced by the above factors tends to lower the depth of penetration, the curves in Fig. 1 were drawn as a best average of the high-lying points. The critical angle θ_c was assumed to be at the last point corresponding to volume diffusion. The sample diffused at the intermediate temperature 593°C did not possess as many pairs of sufficiently columnar grains as the other two. For that reason the upper end of the corresponding curve was drawn not as a best fit to the few points but rather as an interpolation between the other two runs. The difference is entirely within the range of scatter typical at these high angles.

ACTIVATION ENERGIES

An expected feature of the penetration curves is the fact that, although the penetration at 45° changed with increasing temperature only by a relatively small amount (because of the choice of times of diffusion), the relative increase of the volume penetration at $\theta<\theta_c$ is large. However, it is important to note that the critical angle θ_c increased much more rapidly than it would follow from the increase of the rate of volume diffusion alone. In other words the drop of the penetration curve near 45° is increasingly steeper with increasing temperature. This indicates that for angles just above θ_c the activation energies are very low.

The depth of penetration, i.e., the distance between the original interface and an observable concentration, depends on time and temperature. Thus, in order to determine the actual value of the activation energy from experimental data, it is necessary to know the dependence on time or to compare penetrations at various temperatures for a constant time. Most experimental methods make the latter alternative quite impractical. It remains thus to consider a suitable dependence of penetration on time. For volume diffusion we have the usual

$$y_V^2 = Kt \exp(-Q_V/RT), \tag{1}$$

which gives

$$Q_V = R\left(\frac{1}{T_2}-\frac{1}{T_1}\right)^{-1} \log(y_{V1}^2 t_2/y_{V2}^2 t_1) \tag{2}$$

with subscripts 1 and 2 pertaining to two different diffusion experiments. Strictly speaking this formula is valid only for $\theta<\theta_c$ where we obtain $Q_V=34{,}000$ cal/mole. Applying (2) to the whole range of θ gives curve 1 in Fig. 2.

Another possibility is to use Fisher's approximate solution[4] of the grain boundary diffusion equation which is applicable near $\theta=45°$. It leads[5] to

$$y_B^2 = Kt' \exp\left[-\frac{1}{RT}(Q_B - Q_V/2)\right] \tag{3}$$

and

$$2Q_B - Q_V = R\left(\frac{1}{T_2}-\frac{1}{T_1}\right)^{-1} \log(y_{B1}^4 t_2/y_{B2}^4 t_1). \tag{4}$$

Combining Eqs. (1) and (2), we have

$$Q_B = R\left(\frac{1}{T_2}-\frac{1}{T_1}\right)^{-1} \log(y_1 y_{VB1}^2 t_2/y_{V2} y_{B2}^2 t_1). \tag{5}$$

[4] J. Fisher, J. Appl. Phys. **22**, 74 (1951).
[5] R. Smoluchowski, Phys. Rev. (to be published).

Knowing Q_V or knowing the volume penetrations from measurements for $\theta<\theta_c$, we obtain the actual grain boundary activation energy $Q_B=24{,}500$ cal/mole. Applying Eq. (5) to the whole range of grain boundary diffusion gives curve 2 in Fig. 2.

It is clear from Fig. 2 that there is a range of angles for which the activation energy is indeed very low. This is true whether we use as the starting point Eq. (1) or Eq. (3) which are applicable at the opposite extremes of θ. It has been suggested that the grain boundary is a uniform layer of high diffusivity which with decreasing θ, gradually approaches the properties of the unperturbed lattice. If this were true, then the activation energy would gradually increase from a minimum value Q_B at $\theta=45°$, where the boundary is almost uniform, to Q_V at small θ. This is clearly not the case.

The sequence of grain boundary structures suggested previously[1] seems to be at least in a qualitative agreement with the observed facts. With increasing θ the density of dislocations increases rapidly until at a certain critical angle the dislocations bunch up into narrow channels of highly distorted lattice. At still higher angles these channels cluster to form flat areas of fit and misfit resembling Mott's island model.[6] The *apparent* activation energy of grain boundary penetration depends on the balance between the loss of the diffusing material into the two grains and the flux of the material along the grain boundary; therefore it depends strongly on the actual structure of the boundary. For a uniform grain boundary (or nearly uniform at $\theta=45°$) there is relatively much less loss into the two grains than for a boundary made up of narrow channels. Thus in the latter case, i.e., near θ_c, the apparent activation energy will be especially low in agreement with experiment. A more quantitative treatment of this model is the subject of another study.[5]

This work has been sponsored by an AEC contract.

[6] N. F. Mott, Proc. Phys. Soc. (London) **60**, 391 (1948).

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

The Use of Radioactive Isotopes in a Study of Evaporation from Thermionic Cathodes*

W. F. LEVERTON AND W. G. SHEPHERD
Department of Electrical Engineering, University of Minnesota, Minneapolis, Minnesota
(Received April 18, 1952)

Radioactive tracers have been used in a study of evaporation of the alkaline earths, barium, strontium, and calcium from oxide coatings on passive nickel bases. Several percent of the total cathode Ba, Sr, and Ca leave a triple oxide cathode during processing and aging. The rates of transfer of these elements from the cathode during life tests of 118 hours at true cathode temperatures from 1175 to 1279°K obey equations of the form, $\log W=A-(B/T)$, where W is the rate of transfer in g/cm²/sec, T is the absolute temperature of the cathode, A and B are constants. The transfer of Ba from a BaO cathode obeys a similar law. The mass of Ba transferred in 118 hours of life from a triple oxide cathode at 1123°K true as 2.9×10^{-6}g/cm², and from a BaO cathode 1.4×10^{-5}g/cm². The masses of Sr and Ca transferred from a triple oxide cathode at the same temperature were 2.1×10^{-7} and 1.2×10^{-8}g/cm², respectively.

Bombardment of the anode by an electron beam during life tests decreased the quantity of Sr observed on the anode. Anode current densities of 5 to 10 ma/cm² at electron energies of approximately 100 volts produced a fiftyfold reduction in the quantity of Sr on the anode. Experiments indicated that this decrease was not due to heating of the anode. The quantity of Ba deposited on the anode was not decreased by electron bombardment.

The very considerable transfer of Ba from the cathode during processing was greatly reduced by breaking down the carbonates and high temperature flashing the cathodes in the presence of a few mm pressure of argon gas.

I. INTRODUCTION

IN the past, numerous attempts have been made to design vacuum tubes employing the well-known phenomenon of secondary emission. In principle it is possible to build such tubes with a ratio of mutual conductance to input capacitance which is much higher than that obtained in the best conventional tubes available.[1] In general it has been found that the secondary emission ratio of the multiplying surface decreases gradually with life, making stable operation over extended periods impossible. The change in secondary emission ratio with life has generally been attributed to two causes:

(1) changes in the surface caused by electron bombardment,[2] and

(2) poisoning of the multiplying surface by deposition of material evaporated from the primary cathode.[1]

By suitable processing of a silver magnesium alloy

* Supported by the U. S. Army Signal Corps Contract DA 36-039 sc-159.

[1] J. L. H. Jonker and A. J. W. M. van Overbeck, Wireless Engr. **15**, 150 (1938).

[2] C. S. Bull and A. H. Atherton, Proc. Inst. Radio Engrs. **97**, 65 (1950).

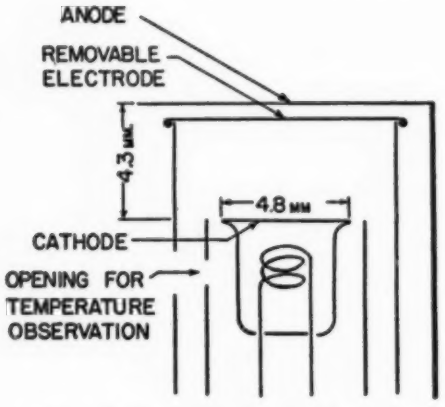


FIG. 1. Schematic diagram of test diode.

Zworykin, Ruedy, and Pike[3] were able to produce a secondary emitter which was stable under moderate electron bombardment for at least 1,000 hours of life. Poisoning by cathode material can be avoided to some extent by employing a tube structure in which the secondary emitting surface is not exposed to the primary cathode.[1] This result is achieved only by some sacrifice of close spacing and simplicity of tube design.

A program has been undertaken in this laboratory to study secondary emission with the particular objective of obtaining a secondary emitting surface having desirable characteristics when exposed to a hot oxide-coated cathode. A quantitative knowledge of the rates of transfer of barium, strontium, and calcium from oxide-coated cathodes for various processing schedules and operating conditions would be very useful in such a study. Knowledge of these rates is meager. Results obtained by chemical methods are subject to large errors because the quantities transferred at typical cathode temperatures are extremely small. Debiesse[4] added traces of radioactive Ba_{140}, Sr_{89}, and Sr_{90} to a conventional oxide coating in commercial pentodes. He found that comparable amounts of barium and strontium left the cathode during processing, but he was unable to make any quantitative measurements of the rate of transfer during life tests.

This paper describes results obtained by the radioactive tracer method. Radioactive Ba_{140}, Sr_{89}, and Ca_{45} have been incorporated in cathode coatings and their rates of transfer measured for various processing schedules and operating conditions during life. A special tube structure permitted distinguishing between material transferred during processing and that transferred during life.

II. METHOD

One to ten mc of one of the isotopes as the carbonate was prepared in one gram of carbonate mixture containing the desired proportions of $BaCO_3$, $SrCO_3$, and $CaCO_3$. The carbonate mixture was ball milled with a nitrocellulose binder and painted on the cathode base by means of an eye dropper.

The radioactive cathodes were used in simple diodes (Fig. 1). In order to differentiate between material evaporated during reduction and activation of the cathode and that evaporated during life, a removable electrode completely hid the anode from the cathode during processing. After activation, tilting the diode allowed the removable electrode to drop out of position. During life test the anode was exposed to the cathode. Upon completion of the life test, the diode was opened and the radioactivity of the cathode, anode, and removable electrode was measured and compared with the initial activity of the cathode.

Ba_{140} with a half-life of 300 hours decays to La_{140} with a half-life of 40 hours. If the total activity of a sample originally composed of Ba_{140} is measured as a function of time, the activity increases initially due to the build-up of the daughter product. After a time equal to a few half-lives of the daughter product, the relative concentration of Ba_{140} and La_{140} will have reached equilibrium and the rate of decay of total activity will be defined by the half-life of Ba_{140}.

In the measurements described here the quantity of barium transferred to the anode was determined by measuring the total activity of the anode deposit and comparing to the activity of a known mass of cathode coating. In order to measure accurately the quantity of barium transferred, it was necessary to allow the tube parts to age for approximately one week after the life test was completed. This allowed the relative concentrations of Ba_{140} and La_{140} to come to equilibrium. Thereafter the mass of barium on the anode could be determined by a direct comparison of its activity with that of a known mass of cathode

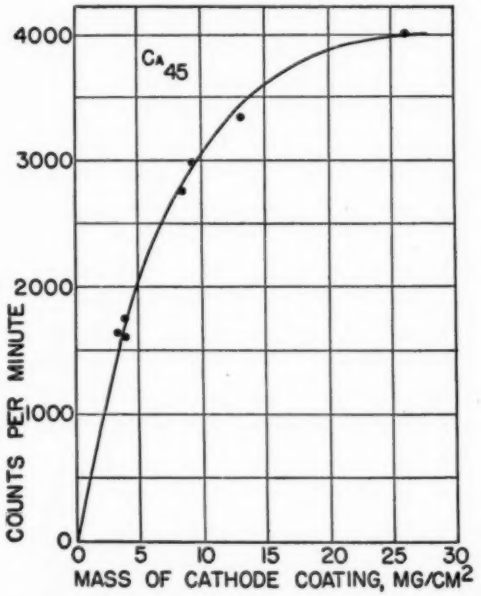


FIG. 2. Radioactivity of triple carbonate cathode coating containing Ca_{45} *vs* coating thickness.

[3] Zworykin, Ruedy, and Pike, J. Appl. Phys. **12**, 696 (1941).
[4] Jean Debiesse, L'Onde Electrique **30**, 351 (1950).

coating. Since Sr_{89} and Ca_{45} decay directly to stable isotopes, quantitative determination of strontium and calcium could be made immediately upon completion of the life test.

The cathode bases were disks of No. 499 (high purity) nickel 4.8 mm in diameter heated by a tungsten filament. The cathode temperature was determined by measuring the brightness temperature of the side of the cathode cup by means of an optical pyrometer. In test diodes this temperature was found to agree in a satisfactory manner with the apparent temperature of the uncoated cathode base as viewed through a 2-mm hole in the anode.

In order to use the radioactivity as a measure of the quantity of coating on the cathodes the apparent activity must be corrected for self-absorption. To obtain this correction a number of sample cathodes were coated and their activity measured. The mass of the coating on each sample was determined by weighing and the activity was plotted against the mass of coating per cm^2 (Fig. 2). The self-absorption correction for Ca_{45} in the useful range of coating densities (from about 2 to 7 mg/cm^2) is not large.

Similar self-absorption curves were plotted for coating containing Sr_{89} and others containing Ba_{140}. The corrections were smaller than that for Ca_{45}.

III. TUBE PROCESSING AND ACTIVATION

The diodes were mounted in Pyrex bulbs, evacuated, and then baked out in vacuum for 1 hour at 450°C. The residual gas pressure was about 10^{-7} mm Hg as indicated by an ionization gauge.

To reduce the carbonates, the filament voltage was raised by steps of 0.5 volt to 6.0 volts (corresponding to a cathode temperature of approximately 950°C). During this process, the cathode temperature was raised slowly so that the gas pressure never exceeded 3×10^{-5} mm Hg. The cathode remained at 950°C until the gas pressure had dropped to approximately 10^{-7} mm Hg (about 30 minutes). At this time the cathode temperature was raised to 1050–1100°C for 2 minutes. The cathode temperature was reduced to 950°C and a positive voltage (125 to 150 volts) was applied to the removable electrode to outgas it by electron bombardment. Approximately 1.2 watts was supplied to the removable electrode for 3 minutes which was sufficient to heat it to a dull red (600–650°C).

A batalum getter was included in each diode. After outgassing the removable electrode, the getter was outgassed, and when the pressure had fallen to 10^{-7} mm Hg it was fully flashed and the diode sealed off. The cathode temperature was kept at 950°C during the outgassing and flashing of the getter.

After seal-off the cathode temperature was set at 850°C. Over a period of 30 minutes the positive voltage applied to the removable electrode was gradually increased from zero until an electron current of from 4 to 6 ma was drawn (approximately 100 volts on anode). The diode was operated under these conditions for 5 hours, the removable electrode was taken out of position, and the diode was ready for a life test.

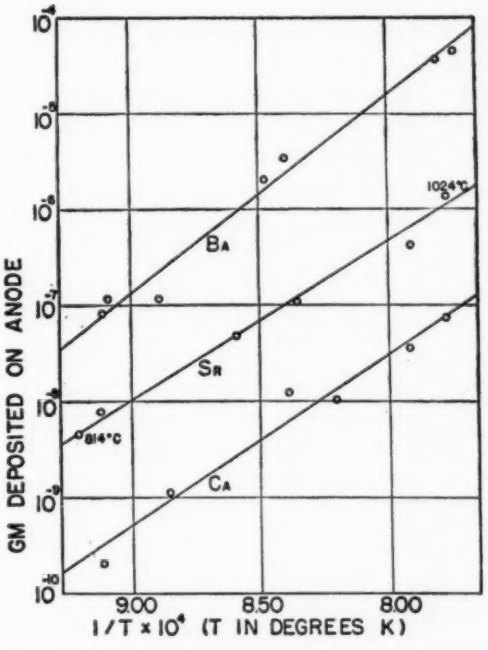


FIG. 3. Masses of Ba, Sr, and Ca deposited on anode during 118 hours of life *vs* cathode temperature for No. 499 nickel cathode bases coated with a triple carbonate cathode coating.

IV. RESULTS

A. Triple Carbonate

For the first group of experiments a triple carbonate mixture of the following composition was used:

$BaCO_3$	49 percent by weight
$SrCO_3$	44
$CaCO_3$	7

(1) *Quantity transferred during processing and aging.*
(a) During processing and aging on the average 1.5 percent of the cathode Ba was deposited on the removable electrode (equivalent to about 80 atomic layers of Ba).

(b) Sr: 1.8 percent (80 atomic layers).
(c) Ca: 0.8 percent (5 atomic layers).

(2) *Quantity transferred as a function of cathode temperature during life.*—A number of diodes were operated for 118 hours with anode currents of about 1 ma (current density 5 to 10 ma/cm^2) and cathode temperatures ranging from 760 to 950°C.

(a) The quantity of Ba deposited on the anode ranged from 5×10^{-2} to 20 percent of the total cathode barium.

(b) Sr: 10^{-3} to 1 percent of the total.
(c) Ca: 10^{-3} to 0.4 percent of the total.

One diode containing radioactive strontium was activated but not life tested. Following activation the removable electrode was taken out of position and the tube was allowed to stand for one week. When the tube was opened and the anode examined for radioactivity none was observed, although amounts as small as 1×10^{-4} percent could have been detected.

In Fig. 3, the masses of Ba, Sr, and Ca deposited on the anode during 118 hours of life are plotted on a log scale against the reciprocal of the true temperature of the cathode. The experimental points are fitted to straight lines by the method of least squares. The geometry of the test diodes was such that 0.73 of the material evaporated from the cathode would be deposited on the anode (or removable electrode).

The quantity of Ba transferred during life tests is about an order of magnitude greater than that of Sr, which in turn is about an order of magnitude greater than the quantity of Ca.

The results given above permit the calculation of an average rate of transfer during the first 118 hours of life. Experiments in which diodes were life tested for periods of from 5 to 350 hours indicated that cathode coating material was transferred to the anode at a greater rate during the first 20 to 50 hours of life than at later stages.

(3) *Effect of emission current.*—(a) Sr: In order to test the effect of emission on the transfer of Sr, a group of diodes were life tested for 118 hours drawing emission currents to the outer shield and anode of from 0 to 6 ma. The temperature of the cathodes was 855°C. The quantity of Sr on the anode decreased with increasing emission current, the quantity deposited being about 50 times greater at zero current than at 6 ma. Log Q plotted against emission current gave a straight line (A in Fig. 4).

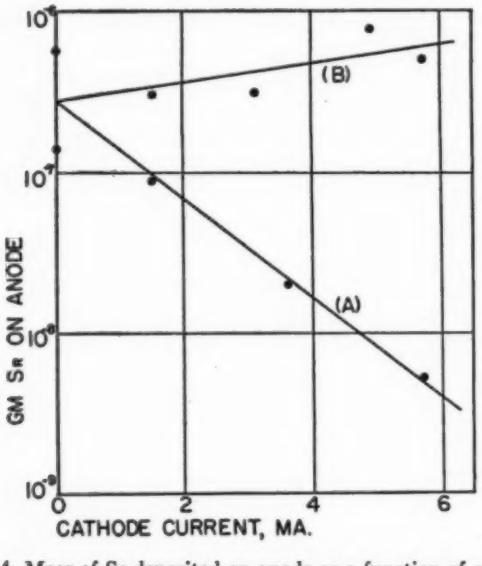


FIG. 4. Mass of Sr deposited on anode as a function of cathode current: (A) Anode bombarded by electron beam. (B) Anode not bombarded.

Another group of diodes was life tested in a similar manner except that the anode was held at minus 45 volts and the entire emission current was drawn to the outer shield. For this group the quantity of Sr transferred to the anode changed little with emission current over the same range of currents, (B in Fig. 4), indicating that the quantity of Sr deposited on the anode is not a function of the cathode current but rather, it is a function of the anode current.

(b) Ba: A number of diodes containing radioactive barium were life tested at 855°C drawing emission currents of from 0 to 6 ma to the anode and outer shield. The quantity of Ba on the anode did not vary significantly with emission current.

The rate of transfer of Ba at this temperature is 10 to 20 times greater than that for Sr; hence if "stripping" (or re-evaporation) of Ba occurred at a rate similar to that for Sr, the effect would be largely masked by the greater deposition rate. In order to avoid this possibility a second group of diodes was life tested in a similar manner at 770°C, at which temperature Ba deposits at a rate comparable to that for Sr at 850°C. Under these conditions the quantity of Ba on the anode did not vary appreciably with emission current.

B. Single Carbonate of Barium

Diodes were prepared with cathode coatings of pure $BaCO_3$ containing radioactive Ba.

(1) During processing and aging on the average 5.6 percent of the cathode Ba was deposited on the removable electrode.

(2) Diodes were operated for 118 hours at cathode temperatures from 775°C to 910°C. The quantity of Ba deposited on the anode ranged from 0.1 to 30 percent of the total cathode barium.

In Fig. 5 the mass of Ba deposited on the anode is plotted on a logarithmic scale against the reciprocal of the true temperature of the cathode. The results are fitted well by a straight line (curve A). For comparison, curve B for Ba transfer from the triple carbonate is shown. Also, for comparison this curve is redrawn (C) to show the amount of Ba which would evaporate from the triple carbonate cathode if it were increased in area until it contained the same number of Ba atoms as the single carbonate cathode. Finally this figure shows the amount of Ba which would deposit on the anode from a barium oxide source at the cathode temperature in 118 hours (curve D). This curve is calculated from the vapor pressure data of Blewett, Liebhafsky, and Hennelly[5] and of Claassen and Veenemans[6] for barium oxide. The mass deposited from a Ba source would, of course, be many orders of magnitude greater.

[5] Blewett, Liebhafsky, and Hennelly, J. Chem. Phys. 7, 478 (1939).

[6] A. Claassen and C. F. Veenemans, Z. Physik 80, 342 (1933).

C. Single Carbonate of Strontium

(1) During processing and aging of diodes with cathode coatings of pure $SrCO_3$, 0.1 percent of the cathode strontium was deposited on the removable electrode, while others processed and aged at similar cathode temperatures throughout but without drawing emission current had 0.3 percent on the removable electrode.

(2) Preliminary experiments indicate that the quantity of strontium deposited on the anode during life decreases with increasing anode current as was the case for strontium from the triple carbonate.

In an effort to determine to what extent the observed variation with current of the quantity of strontium deposited on the anode might result from re-evaporation caused by bombardment heating of the anode, a number of tubes were life tested with the anodes heated to various temperatures and with no emission current drawn. Within the statistical error no variation in the anode deposition was found for anodes operated at temperatures up to 650°C. In the experiments previously described the bombardment heating of the anodes was sufficient to raise the anode temperatures to a maximum of 400 to 500°C; hence the above result indicates that the variation of anode deposition cannot be due to a simple re-evaporation of strontium from the anode.

D. Use of Inert Gas During Processing Triple Carbonate Cathodes

During processing of triple carbonate cathodes in vacuum, from 1 to 2 percent of the total barium, strontium, and calcium is transferred to the anode. This corresponds to a total of the order of 200 atomic layers on the anode, a considerable part of which is deposited during the reduction of the carbonates and the high temperature forming.

In an effort to reduce this transfer, a number of diodes containing radioactive barium were processed with a few mm pressure of argon gas present during these steps. The presence of argon reduced the amount of barium transferred during reduction of the carbonates and high temperature forming by a factor of the order of 100. Another lot of diodes were given the same treatment in argon, and after the high temperature forming the argon was pumped out and the remainder of the processing was performed in high vacuum. These diodes were divided into two groups; in one the outgassing of the anode was accomplished by electron bombardment, in the other the anode was outgassed by heating with rf. Both groups were aged as previously described. The argon treatment reduced the total quantity of barium transferred during the entire processing by a factor of 6 for diodes outgassed by electron bombardment, and by a factor of 20 for those outgassed by rf heating. The additional decrease in barium transfer observed when rf heating was substituted for electron bombardment was presumably due to the fact that when using rf to outgas the anode, the cathode could be operated at a comparatively low temperature. Limited life tests indicated that the rate of transfer of barium during life from cathodes processed in argon was comparable with that of cathodes processed entirely in vacuum.

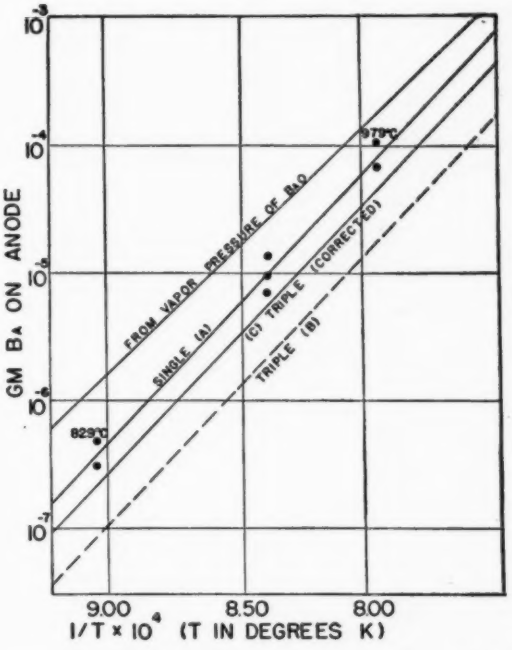


FIG. 5. Mass of Ba deposited on anode during 118 hours of life *vs* cathode temperature for No. 499 cathode bases, (A) coated with barium carbonate and (B) coated with triple carbonate. (C) indicates the mass of Ba which would be deposited from a triple carbonate cathode if it were increased in area until it contained the same total number of Ba atoms as the barium carbonate cathode. (D) indicates the mass of Ba which would be deposited from a barium oxide source at the cathode temperature in 118 hours.

The over-all reduction of Ba transfer by a factor of approximately 20 represents a significant advance in the direction of reducing poisoning of other electrodes by the cathode. In particular, the use of an inert gas during processing might reduce grid emission in conventional tubes.

The chief remaining source of transfer of Ba during processing is the aging of the cathode. A further reduction may be achieved by aging at lower temperatures.

E. Purity of Experimental Diodes

In a spectrochemical study of cathode impurities, Nottingham, Cardell, and Levy[7] found evidence that impurities will migrate to the cathode from other tube parts such as the cathode tabs. The test diodes used in the experiments reported here had cathode disks and cups of No. 499 nickel, but the cathode support tabs were of nilvar and the heat shield and anode were of

[7] Nottingham, Cardell, Levy, Summary Report on ONR Contract N7onr-389 (July, 1950).

TABLE I. Constants for evaporation of cathode coating elements.

Cathode coating	Element	A	B
Triple carbonate	Ba	7.43	2.10×10^4
Triple carbonate	Sr	2.40	1.66×10^4
Triple carbonate	Ca	2.21	1.78×10^4
Barium carbonate	Ba	8.38	2.13×10^4

Grade A nickel. In order to determine whether the presence of these less pure tube parts has influenced the activation and rate of transfer of barium, strontium, and calcium, it is planned to repeat some of the experiments using diodes in which the entire assembly, including the anode, is made of No. 499 nickel.

Preliminary experiments indicate that the rate of transfer of barium from the single carbonate in "pure" No. 499 nickel diodes may be somewhat higher than in our regular diodes.

V. SUMMARY

The rates of transfer of barium, strontium, and calcium from No. 499 nickel cathode bases coated with a triple carbonate mixture have been determined for cathode temperatures from 1175°K to 1279°K. The rate of transfer of barium from cathode coatings of barium carbonate was also measured. The results are fitted by equations of the form:

$$\log W = A - (B/T),$$

where W is the average rate of transfer of the particular element from the cathode in grams/cm²/sec during the first 118 hours of life,† T is the true temperature of the cathode in degrees K, A and B are constants. The results are summarized in Table I.

The rate of transfer of cathode strontium to the anode was found to decrease exponentially with increasing anode current for anode current densities up to 10 ma/cm². If the anode was held negative and the same total cathode current drawn to a second electrode, the total deposition of Sr on the anode was the same as if no cathode current was drawn. Experiments involving indirectly heated anodes indicated that the apparent decrease with anode current of the deposited Sr could not be attributed to re-evaporation due to general heating of the anode by electron bombardment.

The experiments described here do not provide a clear-cut explanation for the variation of Sr with anode current. It has been suggested that the effect could arise if an appreciable fraction of the Sr left the cathode either as Sr^+ or SrO^+. The dependence of the deposition on anode current would then be explained as a result of the repulsion of the ions by the positive anode. The fact that no variation was found with the anode biased negatively would be assumed to occur as a result of all the positive ions being drawn to the negative electrode. The evidence concerning the emission of positively charged ions from oxide-coated cathodes is conflicting. Aldrich[8] reports that a minute fraction (<0.01 percent) of the total Sr evaporated from a SrO coating on Pt left as Sr^+. These observations were made at temperatures several hundred degrees higher than those used in these experiments. Plumlee and Smith[9] reported no positive ion emission from a (BaSrCa)O coating on Ni for temperatures as high as 1100°C within the limit of the measurement (10^{-10} amp/cm²). The present experiments indicate that most of the Sr would have to leave the cathode in ionized form if the proposed hypothesis were to explain the effects observed, and this would require a current of the order of 4×10^{-8} amperes/cm².

It is not necessary to depend on these arguments to rule out the proposed explanation. If the Sr left the cathode in ionized form, the ions would be expected to have thermal velocities and a retarding potential of less than a volt would prevent them from reaching the anode, whereas the decrease of deposited Sr varied continuously for voltages as great as 100 volts. Finally, with the anode negative and all current drawn to a shield electrode a potential saddle existed between cathode and anode sufficient to prevent any positive ions emitted from the cathode with thermal velocities from reaching the anode. There appears, therefore, to be sufficient evidence to rule out an explanation of the variation of Sr deposition as due to emission of Sr in ion form from the cathode.

The exponential decrease with current suggests the possibility that neutral Sr emitted from the cathode is ionized by the electron stream between cathode and anode and hence drawn back to the cathode. Such a process would give rise to an exponential dependence on current, but a simple calculation based on measured transfer rates indicates that this process would require an ionization probability several orders of magnitude larger than any probable value. We do not believe that too much significance can be attached to the logarithmic variation over the limited range observed. It should be noted that the anode voltage was simultaneously varied.

The most probable explanation for the effect seems to be that it results from a "stripping" of the anode deposit by the electron beam. Indirect evidence for such an effect has been reported by Jacobs.[10] Plumlee and Smith[9] using mass spectrometric methods found no evidence for such stripping from (BaSrCa)O surfaces which could be separated from the purely thermal effect of the bombardment. The radioactive tracer methods used here are, however, several orders of magnitude more sensitive than the mass spectrometric techniques so that the effect observed may have been below the limit of sensitivity of the mass spectrometer. A further study of this phenomenon is planned.

† For the geometry used, only 0.73 of the total material evaporated from the cathode was deposited on the anode. This correction was applied in calculating W.

[8] L. T. Aldrich, J. Appl. Phys. **22**, 1168 (1951).
[9] R. H. Plumlee and L. P. Smith, J. Appl. Phys. **21**, 811 (1950).
[10] Harold Jacobs, J. Appl. Phys. **17**, 596 (1946).

No "stripping" of Ba was observed.

In spite of their widely different rates of transfer from the cathode during life tests, it was found that barium and strontium left the cathode in nearly equal amounts during processing and aging of triple carbonate cathodes. For barium, strontium, and calcium the quantities deposited on the removable anode were 1.5, 1.8, and 0.8 percent, respectively, of the total available in the cathode.

The presence of a pressure of several mm of argon gas during breakdown of the cathode coating carbonates and the high temperature forming followed by outgassing of the tube structure by rf heating in vacuum resulted in a decrease by a factor of 20 in the quantity of barium leaving the cathode during processing and aging. For the geometry used in these experiments this represented a decrease from 80 atomic layers to approximately 4 atomic layers of barium deposited on the removable anode. The use of inert gas during tube processing might considerably reduce grid emission in conventional tubes and should reduce the poisoning of the sensitive target surface in tubes employing secondary emission.

The radioactive tracer method is a convenient and accurate method for determining rates of transfer of cathode coating materials under various conditions. The method does not distinguish between free metals and their oxides (or other compounds) but measures only the total quantity of the element transferred. The authors have, for convenience, referred to quantities of barium, strontium, and calcium transferred but do not wish to imply that these elements leave the cathode primarily as free metals. There is evidence from mass spectrographic studies[8,9] that the alkaline earths leave the cathode, to a considerable extent, in the form of oxides.

VI. ACKNOWLEDGMENT

We wish to acknowledge the assistance of P. Wargo, R. W. Peterson, and W. L. Guderian in carrying on this work. We are indebted to Dr. A. van der Ziel for many helpful discussions.

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

The Reliability of Internal Standards for Calibrating Electron Microscopes*

JOHN H. L. WATSON,
Edsel B. Ford Institute for Medical Research, Henry Ford Hospital, Detroit, Michigan
AND
WILLIAM L. GRUBE,
Research Laboratories Division, General Motors Corporation, Detroit, Michigan

(Received March 31, 1952)

Conclusions important to the practice of electron microscopy are drawn from experiments. Particles of Dow Latex 580G are unstable under electron beams and best micrographed at low intensity. Two different effects of radiation are noted. At low intensity the mean diameter in the present work was 2890A, the standard deviation of the spherical particles was 60A. 580G is improved as an internal standard if recalibrated by some other suitable standard (not the reverse) each time before use. Fresh replicas of diffraction gratings are still the most practical standards, reproducing gratings to two percent. By an independent optical method (spectroscopy) one can measure the grating space of a replica when it is on an eighth-inch, 200-mesh specimen screen under exact calibration conditions for electron microscopes. Specific replicas, and from them specific preparations of 580G are checked easily and accurately by the spectroscopic method. Replica age or other factors which may affect accurate reproduction of an original grating are then no longer important in calibration. A reliable method of calibration using 580G is suggested.

INTRODUCTION

WHILE checking a number of latex emulsions (used in light scattering experiments) for particle size by the procedures of electron microscopy, consistent large discrepancies were observed between these determinations made by us and those quoted by others for the same materials. For Dow Latex 580G, which is specifically of most interest to electron microscopists, instead of the 2590±25A found originally[1,2] a value in the neighborhood of 2890A was obtained—a difference of 11.6 percent. Similar differences were observed with the other latex samples. Such a large deviation was not immediately explainable and since it was at odds with the results of several other laboratories[3,4] (in agreement with some) and since many other groups were sufficiently confident of the 2590A value and of its invariability to use it as an internal standard for all of their microscope work, it was deemed worthwhile to check carefully the possible sources of the differences. If there was an error in measurement the main factors which could be involved were ex-

* Presented in part before the annual meeting of the Electron Microscope Society of America, Philadelphia, Pennsylvania 1951.

[1] R. C. Backus and R. C. Williams, J. Appl. Phys. **19**, 1186 (1948).

[2] R. C. Backus and R. C. Williams, J. Appl. Phys. **20**, 224 (1949).

[3] Charles H. Gerould, J. Appl. Phys. **21**, 183 (1950).

[4] K. L. Yudowitch, J. Appl. Phys. **22**, 214 (1951).

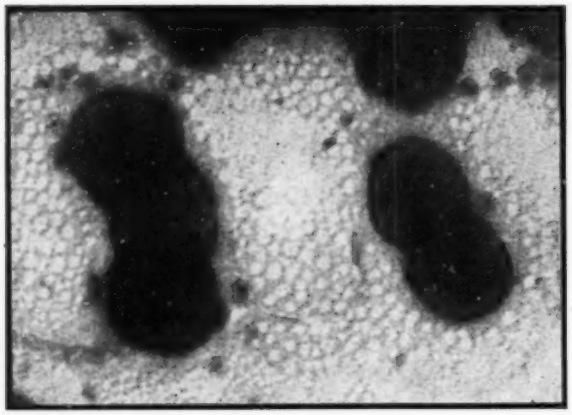


FIG. 1. Dow Latex 580G, after a sudden intense electron bombardment without any prior exposure to an electron beam. ×50,000.

pected to be (1) possible changes in the latex particles themselves while standing in suspension during specimen preparation or during electron microscopic examination and (2) the accuracy of instrument calibration.

During the work, information was gathered on certain properties of the latices themselves as well as on methods of electron microscope calibration, including the use of the 580G latex, glass fibers, and replicas of diffraction gratings as calibration materials. It is this information which is offered in the present paper.

Three stock sources of 580G supply were used: Edsel B. Ford Institute for Medical Research, General Motors Research Laboratories Division, and Wayne University Chemistry Department. With negligible variation all of these showed the same elevated value of the mean diameter. The latices were mounted on a variety of substrates, Formvar, silica, and chromium-covered Formvar, but in no case were significantly different readings recorded for the mean size. Samples diluted from the original concentrated suspensions and mounted immediately were of a size identical with water dilutions which had stood for a variety of periods before preparation.

EFFECT OF ELECTRON BOMBARDMENT UPON 580G LATEX

It was of interest to discover if electron bombardment affected the particles in any way to account for the discrepancies. Two interesting phenomena were observed. First, if the particles of the emulsion were subjected to intense electron bombardment *without any prior exposure to the beam*, they changed instantaneously to the appearance in Fig. 1 wherein there has been a change in contrast so that the spheres have taken on a "hollow" appearance. They often suffered some deformation in their shapes with or without deposit of material about the particles on the supporting substrate, but only minor changes in diameter (usually slight increases). Further bombardment even at high intensity gave rise to no additional changes. To explain the observations it is assumed that volatile substance is driven off from within the particle without noticeable diminution in particle dimensions, so that the density is suddenly lowered.

A surface layer of considerably greater scattering power than the internal volume is observed in these particles (Fig. 1) and contributes to their "hollow" appearances. The thickness ranges from 115 to 178A. Its origin might be in a layer of emulsifying agent or other organic material surrounding the particles, polymerized by the original low intensity electron beam. The observation supports the hypothesis of Ellis† to account for variations in their absolute size that the particles of 580G are covered with a surface layer.

The second effect of bombardment is observed when the field is subjected to a low intensity electron beam first in the process of micrographing it, and is subsequently intensely illuminated. In these cases the phenomenon described above does not occur. Instead there is an immediate *shrinkage* of the particles with no loss of and probably an increase in contrast as in Fig. 2 and no indication of a surface layer. The initial shrinkage is usually quite marked and appears to be an inverse function of the amount of "fixing" action of the low intensity electron beam. Figure 3 shows three typical curves illustrating what happened to the diameters of several individual particles with time of prolonged bombardment. There is a rapid initial shrinkage in a matter of seconds, a slow decrease in diameter for a period of 3 minutes or so, and then a slow increase due to contamination. The increase due to contamination is a very low percentage of the total change and in these experiments it is concluded that any change in particle size due to contamination is negligible.

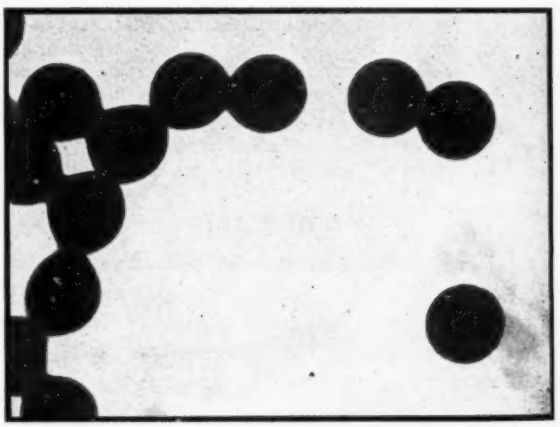


FIG. 2. Dow Latex 580G, after an intense electron bombardment where there had been initial low intensity radiation incident upon the field. ×50,000.

† From unpublished research and observations reported by Ellis to the 1951 meeting of the Electron Microscope Society of America.

In Fig. 3 the upper curve represents a run for which the change was a minimum; the lowest curve one where the change was a maximum; the center curve is drawn for the mean points of all the runs. In the largest change recorded a particle of original diameter 3240A was driven down to 2570A. Probable volatile material is driven off again in this procedure, but the concurrent shrinkage is more than enough to keep the density and the contrast almost constant.

In addition to explanations involving a shrinkage of the particles, these observations might be explainable by assuming again a surface layer, driven off more noticeably at first, percentage-wise, and then slowly as the amount available for evaporation decreases or as the remainder solidifies under the action of the electron beam.

CALIBRATION WITH A GLASS FIBER

Calibration with a glass fiber depends on two operations: measurement of the fiber with a light microscope and micrographing it in the electron microscope. The inherent difficulties and errors of the first operation lead to inaccuracies which are well known and need not be discussed here. The second operation is difficult because of charging phenomena[5] upon relatively large objects under the electron beam.

CALIBRATION WITH A REPLICA OF A DIFFRACTION GRATING

Before any conclusions concerning 580G latex as an internal standard for electron microscopes can be drawn from these observations it is necessary that the measurements be well calibrated. For this purpose a series of Formvar replicas from two 15,000 lines per inch gratings were used. The one series were made by stripping the Formvar replica with Scotch tape[6] and

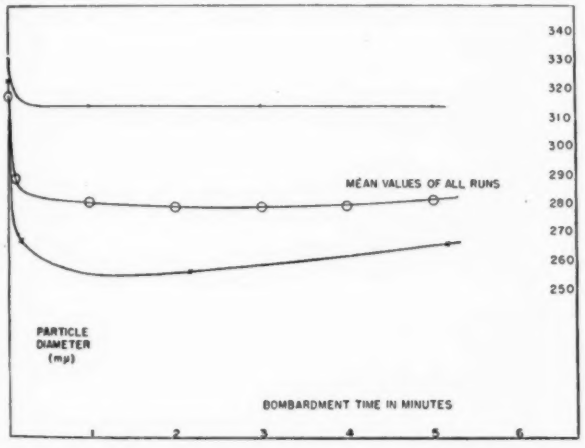


FIG. 3. Variation of the particle diameter of 580G Dow Latex Emulsion with prolonged intense electron bombardment. The upper curve is for a sphere which changed little, the bottom curve is for one which varied considerably. The center curve represents the mean values of all the runs.

[5] John H. L. Watson, J. Appl. Phys. **22**, 1387 (1951).
[6] V. J. Schaefer, Science **97**, 188 (1943).

TABLE I. Grating: Baird Associates, 15,000 lines/inch. Calculated spacing: 1.693 microns. Conditions of measurement: reflection spectrum, Gaertner spectroscope.

Grating	Measured grating spaces (microns) Na spectrum	Hg spectrum
Original grating		
1. G.M.	1.696	1.698
2. E.B.F.I.		1.696
Replica—backed with parlodion and mounted on glass (unshadowed)	1.680	
Replica—backed with parlodion and mounted on glass		
1. chromium shadowed @ 90°	1.684	
2. chromium shadowed @ 27°	1.679	
Replica—backed with parlodion and mounted on large piece of screen		
1. 100 mesh chromium shadowed @ 90°	1.644[a]	
2. 200 mesh chromium shadowed @ 90°	1.678[a]	
3. 100 mesh chromium shadowed @ 90°	1.675[b]	

[a] Poor spectrum obtained. Difficult to make accurate measurements.
[b] Transmission spectrum with Hg source. Better spectrum obtained and therefore a more accurate measurement.

the second series were made, using a modified "backing" method.[7] The replicas were shadow cast with chromium at angles shown in Table I. At least ten fields were micrographed for any one calibration, and an average of the ten "*d*-values" was used in calculating the magnification. No significant discrepancies were noted in the work either between the two gratings or between the two methods of preparing the replicas.

A CHECK OF THE REPLICAS OF A DIFFRACTION GRATING BY SPECTROSCOPIC MEANS

It was decided to check the reliability of the calibration by independent methods. Usual light microscopy and phase microscopy were tried first. To the limits of accuracy of the method the *d*-values of the replicas reproduced 15,000 lines per inch, but because of the inherent difficulties and inaccuracy of the method (the same criticisms would hold relative to the use of a light microscope to calibrate using any macroscopic object, diatoms, glass fibers, etc.) recourse was had further to measuring the *d*-values of the gratings and their replicas by spectroscopic means. It was planned to measure the replicas after they had been subject to a variety of conditions and particularly to conditions similar to those experienced in the electron microscope itself.

The instrument used to measure the *d*-values of the gratings (both original and replicas) is a standard Gaertner student spectrometer. After aligning the spectrometer using an accepted method[8] or by following the instructions supplies by the manufacturer, the gratings were placed in the optical system as the dis-

[7] W. L. Grube and S. R. Rouze, A.S.T.M. Bulletin No. 179 (1952).
[8] G. S. Monk, *Light Principles and Experiments* (McGraw-Hill Book Company, Inc., New York, 1937), Appendix IV, p. 426.

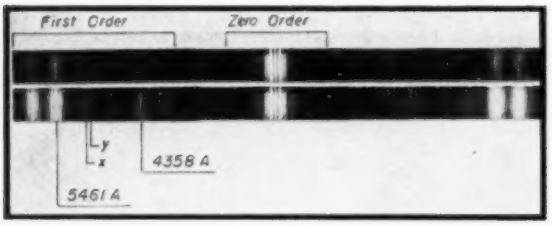


FIG. 4. Typical spectrograms of a mercury source diffracted by a shadow-cast, Formvar replica of a diffraction grating (15,000 lines/inch) mounted upon an $\frac{1}{8}$-inch electron microscope specimen screen.

persing element, and the spacing determined, using the relationship

$$d=n\lambda/[\sin i+\sin\theta],$$

where i=angle of incident beam, θ=angle of diffracted order "n," and λ=wavelength.

The original gratings were measured by reflection using both the sodium and mercury spectra. The d-value determined for the General Motors' grating was 1.696 microns using the sodium spectrum and 1.698 microns using the mercury spectrum. The d-value of the Edsel B. Ford Institute grating determined with the mercury spectrum was 1.696 microns. These values compare favorably with the calculated value of 1.693 microns for a 15,000 line per inch grating (Table I).

Although it was planned to make spectroscopic measurements eventually on replicas mounted on actual electron microscope specimen screens which had been subjected to dessication, heating, and electron bombardment in the microscope, it was decided to measure first replicas which were better mounted for making dispersion studies. To this end Formvar replicas were prepared and measured as indicated in Table I.

The entire replica of the original grating when mounted on glass was found to be of sufficient quality to be measured accurately by reflection diffraction with the sodium spectrum, but when mounted upon mesh poorer reflectivity necessitated use of transmission methods with a mercury source. In Table I it can be seen that wherever accurate measurements were possible the gratings themselves and the replicas from them were always within one percent of the value 1.693 microns, and that where less accuracy is to be expected, e.g., when the replica is mounted on screen, values are within one or two percent of 1.693 microns.

The next step involved the measurement of the shadow-cast replicas when mounted on the electron microscope specimen screens. When prepared on these round, one-eighth-inch, 200-mesh screens and used as transmission gratings with a mercury source, the replicas gave a first-order spectrum of sufficient intensity to be measured. *Thus it was possible in practice to make an independent determination of the grating space on the actual replica used to calibrate the electron microscope.* Figure 4 shows typical spectrograms obtained from a shadow-cast Formvar replica. Each line of the spectrum is actually a series of lines formed by the superposition of the zero-order diffraction pattern which arises from the periodicity of the screen. This complexity sometimes introduces some difficulty in choosing the actual line which should be measured, but when the entire series associated with a given spectral line is well resolved and is sufficiently removed from neighboring groups, the central line can be determined simply from geometrical symmetry. However, the entire group may not always be clearly visible in the telescope eyepiece, and that portion which can be seen may not be symmetrical intensity-wise with respect to a central line. Consequently, the group may appear to have two centrally located lines of approximately the same intensity, for example, lines marked X and Y in Fig. 4. Since the more intense of these will lie on the side of the central line nearest the zero order, the line of the central "doublet" farthest from the zero order, line X, is the correct one to measure.

When a replica grating is measured, the total diffraction angle $\psi=\alpha+\beta$ (Fig. 5) between both first-order spectra is measured rather than the diffraction angle θ_1 or θ_2 between the zero and first order. This is done for two reasons. First, readings of both first-order spectra can be taken without readjusting the entrance slit—an operation which is necessary if a reading is taken on the zero-order pattern. Second, errors due to misalignment of the grating with respect to the incident beam are minimized. It can be shown that an error much greater than that actually encountered can be tolerated before there is an appreciable effect of misalignment upon the result. However, considerable care is still exercised in arranging the grating in exactly the same position as that of the mirror used earlier to align the spectrometer. This operation can be accomplished quite easily with a fixture designed for holding the specimen screens.

The relationship used to calculate the d-value when ψ is measured instead of θ_1 or θ_2 for first order is

$$\lambda/d=\sin\tfrac{1}{2}(\alpha+\beta)[\cos i\cdot\cos\tfrac{1}{2}(\alpha-\beta)+\sin i\cdot\sin\tfrac{1}{2}(\alpha-\beta)],$$

which simplifies to

$$\lambda/d=\sin\tfrac{1}{2}(\alpha+\beta)$$

for $i\cong 0$ or $i=0$.

The lines of the mercury spectrum used to measure

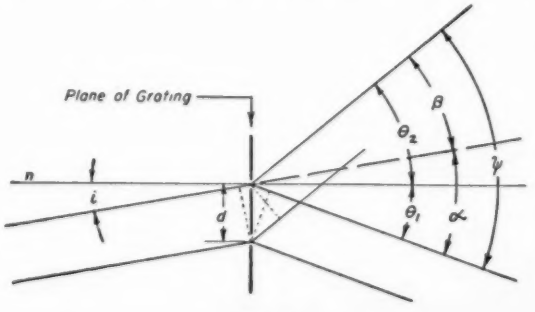


FIG. 5.

the gratings are the 4358A violet and the 5461A green. These lines are indicated on the spectra of Fig. 4.

The *d*-values determined by the method just described are tabulated in Table II for a series of Formvar replicas. Those for the freshly prepared replicas all lie within three percent of the original grating spacing. When the replicas were subjected to dessication in a vacuum, shadow-casting with chromium with its concurrent heating of the replica, and to a short period of electron bombardment, no deviations in *d*-value greater than three percent of 1.693 microns occurred. The grating spaces were also determined for a number of replicas before and after deposition of the latex spheres upon them. Again the *d*-values were found to be well within three percent of the calculated value.

Some old replicas which had been used and had been stored for several months were found to deviate from 1.693 microns by as much as five percent. It therefore appeared that aging might affect the replicas significantly, and it was decided to investigate this important point further. Table III shows that control replicas which had been used and stored for periods as long as 2½ months showed no significant change with aging. Table III also shows the results of spot checks made on aging after 6 months. Again no effect of aging was recorded.

TABLE II. Spectroscopic measurements made on a series of shadow cast grating replicas. Mounted upon one-eighth-inch electron microscope specimen screens.

No.	Date made	Measured spacing (μ)	Date shadowed	Shadow angle	Spacing (μ)	Date 580G added	Spacing (μ)
1	8–10	1.66	8–13	2:1	1.67	[a]	
2	8–10	1.66	8–14	3:1	1.66	8–15	[b]
3	8–10	1.70	8–13	2:1	1.70	8–15	[b]
4	8–10	1.68	8–14	2:1	1.68	[a]	
5	8–10	1.70	8–14	2:1	1.71	8–15	[b]
6	8–10	1.69	8–14	2:1	1.69	8–15	[b]
7	8–10	1.68	8–14	2:1	1.68	[a]	
8	8–10	1.70	8–14	2:1	1.70	8–15	[b]
9	8–13	1.68	8–14	3:1	1.67	[a]	
10	8–13	1.67	8–14	3:1	1.67	8–17	1.67
11	8–13	1.68	8–14	2:1	1.68	8–15	[b]
12	8–13	1.65	8–14	3:1	1.65	8–17	1.66
13	8–13	1.67	8–15	1:1	1.67	8–17	1.67
14	8–13	1.65	8–15	1:1	1.65	8–17	1.67
15	8–13	1.69	8–15	2:1	1.69	8–15	[b]
16	8–13	1.67	8–15	1:1	1.67	[a]	
17	8–13	1.68	8–15	2:1	1.69	8–17	
18	8–13	1.66	8–15	1:1	1.67	[a]	
19	8–22	1.70	8–22	4:1	1.71	8–24	1.71
20	8–22	1.76	8–22	4:1	1.76	8–24	1.79
21	8–22	1.74	8–22	4:1	1.73	8–24	1.74
22	8–22	1.69	8–22	4:1	1.70	8–24	1.70
23	8–22	1.73	8–22	4:1	1.72	8–24	1.73
24	8–21	1.67	8–21	4:1	1.68	8–23	1.70
25	8–21	1.66	8–21	4:1	1.66	8–23	1.67
26	8–21	1.69	8–21	4:1	1.69	8–23	1.69
27	8–21	1.66	8–21	4:1	1.66	8–23	1.67
28	8–21	1.67	8–21	4:1	1.65	8–23	1.67
29	9–19	1.70				9–20	1.71
30	9–19	1.70				9–20	1.71
31	9–19	1.69				9–20	1.69

[a] No. 580G added.
[b] 580G added, too thick. All replicas were Formvar type except numbers 29, 30, and 31, which were SiO_2.

TABLE III. Effect of aging upon replicas of a diffraction grating. (Spectroscopic measurements of the grating distance.)

No.	Spacing [microns] Original	After aging periods 1 month	2½ months	6 months
1	1.67	1.68	1.68	
4	1.68	1.69	1.69	
5	1.71	1.69	1.70	
7	1.68	1.68	1.68	
8	1.70	1.70	1.71	
9	1.67	1.68	1.68	1.68
10	1.67	1.68	1.68	
12	1.66	1.67	1.67	1.69
13	1.67	1.68	1.69	
14	1.67	1.67	1.67	
16	1.67	1.68	1.68	
18	1.67	1.68	1.68	
19	1.71	1.70	1.71	
22	1.70	1.71	1.70	1.71
23	1.73	1.74	1.75	
24	1.70	1.71	1.73	1.73
25	1.67	1.66	1.68	
26	1.69	1.69	1.69	1.69
27	1.67	1.68	1.68	
28	1.67	1.67	1.68	
30	1.71		1.72	
31	1.69		1.71	

DISCUSSION

Since the preparation of the latex specimens seems to introduce no anomalies and since the grating replicas used to calibrate the magnification of micrographs measure well within three percent, it must be concluded that mean diameter of the emulsion particles of 580G measured in this work from three different stocks are greater than 2590A and are of the order of mean diameter 2890A. If it is assumed that the 2590A value was correct when measured, then some type of growth must have occurred in the meantime while the particles stood in suspension or were handled for the electron microscopy. That they are unstable under the electron beam (a process equivalent partially to being heated in a vacuum) has been demonstrated. This instability could be an indication that the particles are able to absorb water on standing, and that their structure "loosens," so to speak, or that they possess a volatile surface layer which is either permanently associated with the particles in suspension or is deposited upon them during drying.

From these studies several conclusions of importance to the practice of electron microscopy can be drawn. (1) The particles of Dow Latex 580G are unstable under intense electron beams and if a reproducible, dependable diameter is required they must be micrographed at low intensity. (2) Although they are both uniform in size in a single preparation and spherical in shape, the uncalibrated 580G spheres are not dependable from specimen to specimen or time to time as far as mean diameter is concerned, having in the present case a mean size of the order of 2890A. (3) Because of points (1) and (2) above, the value of the Dow Latex 580G as an internal standard for calibrating electron microscopes is improved by cali-

brating it each time before use by a suitable standard (not the reverse). (4) Fresh, shadow cast replicas of a diffraction grating are still the most practical and accurate standards for calibrating electron microscopes, and can reasonably be expected to reproduce the grating to two percent (any deviation is generally a shrinkage). (5) It is possible to measure accurately by an independent optical method (spectroscopy) not only the *d*-value of the original grating, but also the grating space in the replica when on a specimen screen under exact conditions necessary for calibration of the electron microscope. (6) Because of point (5) above, one can check a specific replica for its *d*-value conveniently and accurately before using it as a standard for the electron microscope. Age of the replica or other factors which might affect its accurate reproduction of the original grating are then no longer important in a calibration. (7) There is an important corollary to (4) above. Since these plastic replicas reproduce a grating so well, they can be used with equal confidence to replicate any surface and to obtain with adequate accuracy quantitative measurements of solid body properties. The present need for checking electron microscope results against the light microscope or other secondary devices is then obviated, electron microscopy alone of the simple cast replica being sufficient.

A RELIABLE METHOD OF CALIBRATION

The reliability of the shadow cast Formvar replica of a diffraction grating as a standard was checked further by micrographing at random 107 different grating spaces and determining their standard deviation. An average magnification of ×11,400 was secured for the projector lens setting which was chosen and the standard deviation was calculated to be ×330. This is a standard deviation of approximately 3 percent which is of the same order as the variations which occurred in spectrographic measurements made upon the replicas. There is therefore good agreement between the "average" reading obtained with the spectrograph and a statistically accurate mean of individual readings recorded by electron microscopy.

Since it is tedious to measure a statistically accurate number of spaces with the electron microscope and since by this method there can be a range of anywhere from ×10,400 to ×11,800 in magnification (by actual measurement for the 107 cases), it would appear to be practical and desirable to use 580G as an internal standard by depositing it upon a replica calibrated spectroscopically. This can be done without changes occurring in the replica. The size of the 580G particles in the preparation can be estimated accurately against it. This dry preparation can then be stored and used repeatedly with confidence to calibrate the microscope conveniently. New preparations could be made up easily from time to time as earlier ones became contaminated or otherwise un-useable. The uniformity of the particle diameter per preparation regardless of possible changes in mean diameter would necessitate that only a relatively small number of measurements would be needed for a calibration. This done for one setting of projector current, a single particle chosen from the original field and micrographed at each of a series of projector settings would give a rapid calibration of the microscope.

Letters to the Editor

Note on the Strain Aging of Iron Single Crystals

A. N. Holden and F. W. Kunz
Knolls Atomic Power Laboratory, Schenectady, New York*
(Received March 13, 1952)

RECENTLY, several single crystals of high purity iron† in the form of wires several inches long and 0.025-inch in diameter were grown for use in internal friction damping experiments. These crystals were grown under conditions that permitted 0.005 percent carbon to remain in the crystals. Tensile tests performed on a rigid, screw-type machine indicated that these crystals had only very small initial yield points, but were very susceptible to strain aging and exhibited very marked yield point effects after straining and aging short times at room temperature. Elongations in all tests were measured over a one-inch gauge length by cross head motion. The crystals were suspended by six-inch long, 0.025-inch piano wires. This method of suspension gives an erroneous impression of the elastic behavior of the crystals but is quite satisfactory for yielding studies.

Several interesting results of the tests on these crystals are described in this note because it is felt they are pertinent to a theoretical understanding of the strain-aging phenomenon.

Annealed single crystals containing 0.005 percent carbon were, in general, observed to have either very small initial yield points or to have no initial yield point at all. However, after subjecting the crystal to a tensile stress‡ of roughly ⅔ the yield value (of stress) for one hour prior to testing, a marked increase in the upper yield point occurred over that of an unstressed crystal. Such a test is shown in Fig. 1.

Furthermore, if a crystal is deformed in tension and aged while still subjected to a high stress, a higher upper yield stress is observed in a subsequent test than would be observed if the deformed crystal were aged with the external stress relaxed. Such a test for two identical crystals, one aged under stress for 30 minute periods and the other aged for 30 minute periods with the stress relaxed is shown in Fig. 2.

All the crystals tested, even though they showed little or no initial yield point in their tensile tests, had a concave region in the early portion of their flow curve where they were evidently strain hardening at an increasing rather than at a decreasing rate. This observation is apparent in Fig. 1. Such a concave region is characteristic of nearly all impure metal single crystals.

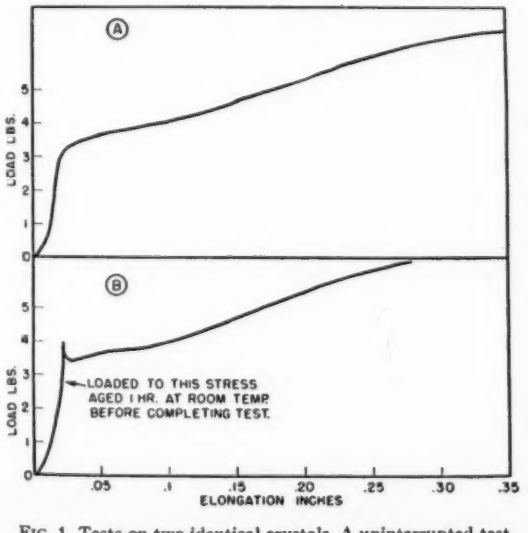


FIG. 1. Tests on two identical crystals, A uninterrupted test, B aged at 2.7 # load for one hour.

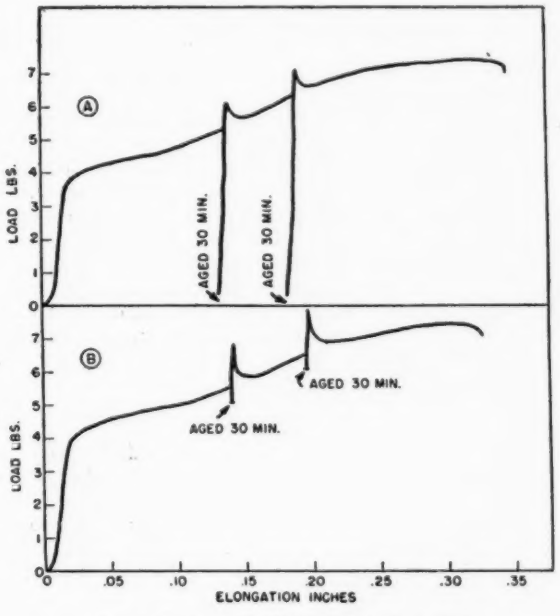


FIG. 2. Strain aging of identical single crystals, A with load relaxed, B with stress applied.

These experiments were done to test a suggestion made by J. C. Fisher[1] that when dislocations are anchored by solute atoms in a "bowed" or stressed condition it is more difficult to pull them free from their solute atmospheres than when dislocations are anchored in their relaxed condition, and consequently, the upper yield point stress is higher when aging is carried out under stress.

* The Knolls Atomic Power Laboratory is operated for the United States AEC by the General Electric Company under Contract No. W-31-109 Eng-52.

† Material obtained from the National Research Corporation and grown into single crystal wires by using a strain-anneal technique. In general, these crystals contain a [110] direction nearly parallel to the wire axis.

‡ It is presumed that no plastic strain occurred because the crystal was tested in a rigid machine and no load drop was observed during the period of stress application.

[1] J. C. Fisher (private communication).

A High Vacuum Flow Analogy

G. W. Monk
Camp Detrick, Frederick, Maryland
(Received March 13, 1952)

IF the Knudsen equation for the mass flow of gas through a tube under high vacuum conditions is written so as to express the net number, N, of molecules per second flowing down the tube as a function of the difference in the number, ν, of molecules per second striking a square centimeter at the two ends of the tube, we have

$$N=\frac{8\pi R^3}{3L}(\nu_1-\nu_2), \tag{1}$$

where R is the radius and L the length of the cylindrical tube.

If Eq. (1) is correct for molecules which travel at fixed speeds, do not interfere with each other, and are emitted from wall surfaces at random angles, it should hold also for the flow of photons down a tube which is rough but highly reflecting. Bright aluminum foil which has been crinkled might provide almost random reflection angles when averaged over an area containing a large number of crinkles. Equation (1) should apply to the flow of electromagnetic energy down a large tube lined with this foil or a similar material if the reflectivity is taken as 100 percent. If each photon is assumed to carry the same amount of energy, the power P con-

ducted through such a tube would be given by

$$P=8\pi R^3(I_1-I_2)/3L, \quad (2)$$

where I_1 and I_2 are the intensities at points a distance L apart in a tube of radius R.

Since the flow of light down a bright tube depends upon the dimensions in the same way that the flow of gas does under high vacuum conditions measurements on the electromagnetic conductance of a simulated high vacuum system should allow determination of the pumping speeds before expensive construction is undertaken. Contemplated traps, baffles, and plenums would be easy to construct of light material lined with foil and alterations could be made to achieve the maximum conductance. Traps for condensible vapors would be highly absorbing but changed to reflecting surfaces to determine the speed for permanent gases in the same system.

The conductance C from Eq. (1) is seen to be

$$C=8\pi R^3/3L, \quad (3)$$

where C is given in cm³. The conductance thus defined is a convenient concept independent of ambient conditions and the nature of the gas, but can be easily converted to more conventional units by multiplying by the quantity of gas striking unit area per second. This factor would be about 11.7 liters/sec cm² for air at room temperature if the pressure is expressed in microns and the mass flow rate in liter-microns per second.

While the reflectivity has been assumed as 100 percent in this discussion, a lower value might not interfere with the usefulness of the method, since further considerations might allow a correction for the absorbtion and in practice relative conductances are generally more important than absolute values. An experimental arrangement to determine electromagnetic conductances might consist of a uniformly illuminated ground glass screen as a source which would allow light to enter a long cylindrical tube in which the rate of decrease of intensity and the dimensions could be used to compute P in Eq. (2). The light would then pass through the simulated high vacuum system and into a black box. Intensities could be measured photoelectrically through small openings in the walls. The loss of intensity through the simulated vacuum system and the known value of P would allow a calculation of C from Eq. (2).

Incidentally, such a system of reflecting ducts could be used to conduct radiant energy in appreciable quantities for practical applications.

Crack Formation on Sodium Chloride Single Crystal Surfaces

ROBERT A. LAD
National Advisory Committee for Aeronautics, Lewis Flight Propulsion Laboratory, Cleveland, Ohio
(Received March 24, 1952)

CRACK formation on the surfaces of sodium chloride single crystals can be studied by a modification of the method used by Andrade and Martendale[1] on glass and quartz surfaces. Deposition of metal films on the crystal surfaces by the usual evaporation techniques at pressures below 10^{-5} mm eliminated the discolorations that they reported occurred when sputtering was used on alkali halides. Electron diffraction photographs of silver films evaporated on sodium chloride show no evidence of reaction between the silver and the substrate.

The sodium chloride single crystal surfaces were prepared by grinding parallel to the (100) face followed by a dissolution of the surface layers in water. Films of silver, gold, and zinc, deposited to a computed thickness of 25 atom layers were observed under a microscope using oblique illumination at an angle of 15° with the horizontal. All three metals were found to deposit as small crystalline aggregates scattered randomly over the surface. On water-treated faces which had been allowed to stand for several months before coating, the aggregates were found to be arranged in straight lines at many angles. Figure 1 is a photomicrograph of such a sodium chloride surface coated with zinc. The fact that the lines occur only after aging suggests that the metal is deposited along cracks in the surface. The same results were obtained by aging the crystals in vacuum or in a desiccator. This precludes the possibility of adsorption of water and subsequent re-crystallization to form steps. In addition, step formation by even such a process

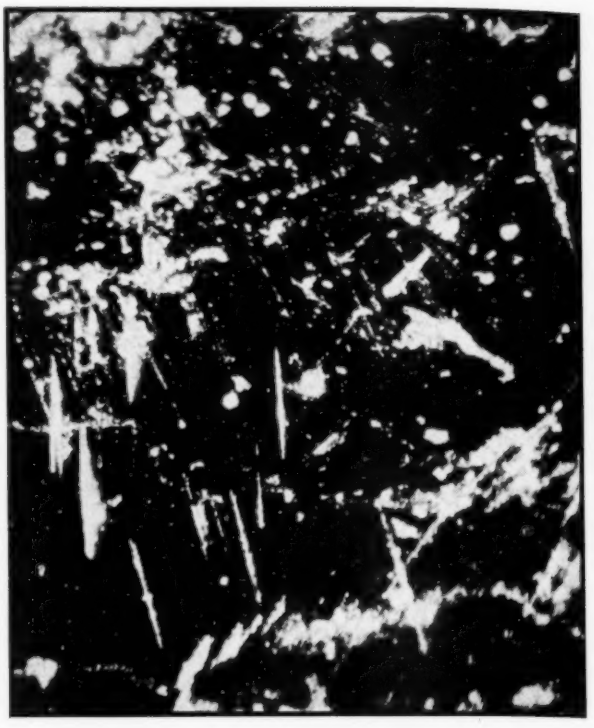


FIG. 1. Zinc crystallites deposited along cracks on the surface of a sodium chloride single crystal (250×).

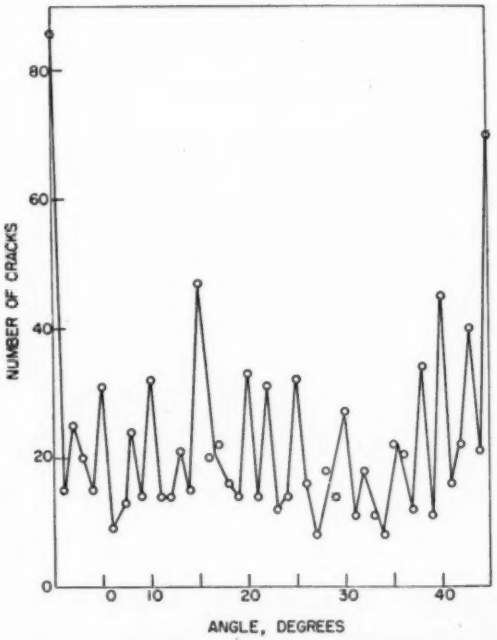


FIG. 2. Frequency of crack occurrence *versus* angle between the crack and a cube edge.

is unlikely since such a process usually results in the formation of small crystals on the surface. Heating of the coated crystals in vacuum caused growth of new crystals in accordance with the observations of others.[1,2] However, the treatment did not result in the formation of more lines, and in some cases the heavy deposit of small crystals obscured the lines which had been visible before heating.

Measurements were made of the angles between the cracks and a cube edge for a number of water treated and aged specimens. The angles were measured to $\pm 0.25°$ and tabulated for one degree intervals. Figure 2 is a typical plot of frequency of occurrence *versus* angle. It can be concluded that the cracks tend to form at certain preferred angles which are the same for all of the specimens. The positions of the maxima in the curves indicate that the tendency to crack in a particular plane is in part governed by the surface energy in the plane.

The deposition of silver films on crystals under tensile stress showed that cracks are formed at stresses which are only a fraction of the tensile strength. Cracks were found on water treated crystals loaded to 0.25 kg/mm². This stress is less than one-tenth of the strength of such a specimen and about one-third of the strength of specimens which had been ground but not water treated. Two hours elapsed between the preparation of the surface and the silver deposition. The cracks on the stressed crystals showed no preference for the direction transverse to the applied stress. The distribution of angles was the same as for cracks formed by aging.

[1] E. N. Andrade and J. G. Martindale, Phil. Trans. **A235**, 69 (1935).
[2] G. W. Johnson, J. Appl. Phys. **21**, 449 (1950).

Note on the Discharge Coefficient of a Critical-Flow Nozzle*

HENRY WISE
Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California
(Received March 28, 1952)

FOR isentropic flow of a perfect gas through a properly shaped nozzle,[1] the rate of mass discharge of gas $\dot{m}$ through the orifice under the conditions of sonic velocity in the throat; i.e., critical flow, is given by

$$\dot{m}=CA_tP(\gamma)^{\frac{1}{2}}[2/(\gamma+1)]^{(\gamma+1)/2(\gamma-1)}/(RT/M)^{\frac{1}{2}}, \quad (1)$$

where A_t represents the geometric throat area of the nozzle; P, the gas pressure on the upstream side of the orifice; γ, the ratio of the specific heat at constant pressure and constant volume at temperature T; R, the gas constant; M, the molecular weight of the gas; and C, the discharge coefficient. Variations in the discharge coefficient given by the ratio of the measured throat area A_t^* to the geometric throat area A_t for a given flow nozzle have been interpreted to be caused by gas impurities.[2] It appears, however, that such an interpretation is not adequate. In the use of a critical flow nozzle with a diameter of $2.02_6\times10^{-2}$ cm for the measurement of chemical equilibria,[3] it was found that the effective throat area varied with molecular weight of the gas as presented in Table I.

TABLE I. Experimental measurement of throat area of nozzle.

Gas	Percent Purity	A_t^* cm²×10⁴	Viscosity η poise×10⁶	$(\eta/M)^{\frac{1}{2}}$ ×10³	$r-r^*$ cm×10²
He	99.8	3.040	199.4[a]	7.07	0.052
N_2	99.9	3.184	184[a]	2.57	0.014
C_2H_4	99.5	3.203	109[a]	1.98	0.010
A	99.9	3.192	221[a]	2.35	0.012
CO_2	99.9	3.202	160[a]	1.90	0.010
CCl_2F_2	97.0	3.220	125[b]	1.00	0.007

[a] Lange, *Handbook of Chemistry* (Handbook Publishers, Inc., Sandusky, Ohio, 1946).
[b] Perry, *Chemical Engineering Handbook* (McGraw-Hill Book Company, Inc., New York, 1950).

It is evident from the data shown that this increase in throat area with molecular weight cannot be ascribed to gas impurities.

On the other hand, the observed effect can be explained by the existence of a boundary layer in the nozzle. Let the effective throat area of the nozzle be given by

$$A_t^*=\pi r^2(1-l/r)^2, \quad (2)$$

where r represents the geometrical radius of the throat, and l, the displacement thickness of the boundary layer.

For a gas stream with pressure and velocity gradients the boundary layer thickness is difficult to evaluate. However, to a first approximation the empirical equation[4] obtained in the absence of such gradients may be applied. Thus for a laminar boundary layer the displacement thickness is given by

$$l=1.73(\eta x/\rho u)^{\frac{1}{2}}, \quad (3)$$

where η is the viscosity; ρ, the gas density; x, the length of the boundary layer; and u, the velocity of the gas.

By introducing Eq. (3) into (2) one therefore obtains for the ratio of the effective to geometric throat area

$$A_t^*/A_t=\{1-(1.73/r)[(\eta RTx)/(uPM)]^{\frac{1}{2}}\}^2. \quad (4)$$

It is apparent from Eq. (4) that for the flow of gases through a nozzle at critical flow conditions this ratio approaches unity with increasing molecular weight and/or decreasing viscosity of the gas at a given temperature and pressure. Additional measurements of the effective throat area for two gases of the same molecular weight but different viscosities such as nitrogen and ethylene (Table I) substantiate the dependency of A_t^* on η as shown in Table I.

Based on Eq. (4) it can further be shown that the difference between the geometric radius r and the measured value r^* is proportional to the square root of η/M. This functional dependency is borne out by the experimental data as shown in columns 5 and 6 of Table I. It is apparent that this boundary layer effect becomes important for nozzles with small orifice area.

* This paper presents the results of one phase of research carried out at the Jet Propulsion Laboratory under U. S. Army Ordnance Department Contract No. DA-04-495-ORD-18.
[1] Lamb, *Hydrodynamics* (Cambridge University Press, London, England, 1932), pp. 23–27.
[2] J. W. Anderson and R. Friedman, Rev. Sci. Instr. **20**, 61 (1949).
[3] H. Wise and D. Altman, J. Chem. Phys. **19**, 515 (1951).
[4] W. F. Durand, *Aerodynamic Theory* (Durand Reprinting Committee, California Institute of Technology, Pasadena, 1943).

On the Redundant Information Supplied in Practical Applications by the Time and Frequency-Phase Responses of a System*

M. LEVY†
Defence Research Board, Ottawa, Canada
(Received March 28, 1952)

MUCH work has been published on methods to determine the step or impulse response of a system from its frequency-phase characteristics and inversely. These are connected by Fourier integrals, and most of the methods are approximate methods of evaluating these integrals. It has not been realized that often in practice these curves contain redundant information which, once eliminated, simplifies considerably the problem. We gave some examples of redundancy in previous studies.[1,2] In this note we shall summarize the problem in its broad but practical aspects, as we presented it at the National Electronics Conference in October, 1951.[3]

Redundancy occurs when we possess some information on the curve which has not been translated by a mathematical operation on it. In practice the most important cases of redundancy are due to the fact that the impulse response does not last indefinitely, or

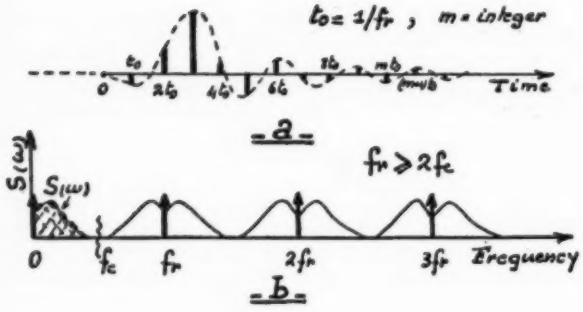


FIG. 1.

that the frequency-phase response is bounded, or that both occur simultaneously.

To study these three main cases in a simple way, we shall consider only systems of the low pass filter type. The high pass and band pass filter types can be reduced to the low pass type, the band pass by considering the envelope of the response instead of the response itself.

First Case: Frequencies above f_c are not transmitted by the system.—This case has already been considered in our previous paper. Multiply the impulse response $I(t)$ by a pulse train of repetition frequency f_r such that $f_r \geq 2f_c$. One obtains a modulated pulse train (Fig. 1(a)) $P(t)I(t)$ whose frequency spectrum is as shown in full lines on Fig. 1(b), the shaded area representing the frequency spectrum of the continuous impulse response. It is obvious that the modulated pulse train contains all the information inherent in the continuous impulse response, since this response can be obtained by sending the modulated pulse train through a low pass filter of cut-off frequency f_c. *Thus, the continuous impulse response contains much redundant information.*

By adding the frequency components of each impulse of Fig. 1a, with the notations of Fig. 1, the frequency phase characteristic is obtained:

$$S(\omega)e^{i\varphi(\omega)} = \text{Re.}\ \Sigma_m\, I(mt_0)e^{-imt_0}. \qquad (1)$$

This equation eliminates all redundancies of the usual Fourier integral. It can be computed graphically in a very simple way.

Second Case: The impulse response vanishes for $t \geq t_c$.—If we assume that the response starts for $t=0$, we can consider the response as one period of a periodic curve (Fig. 2(a)). In this case the frequency-phase characteristic is represented by a fundamental component of frequency $1/t_c$ and harmonic components, while that of the impulse response is the continuous frequency-phase characteristic (in Fig. 2(b) only the frequency characteristic is represented). *It is obvious that this continuous curve contains redundant information.*

From Fourier series equations, or by adding the frequency components of Fig. 2b at time t, with the notations of Fig. 2, we get the impulse response:

$$I(t) = \text{Re}\ \Sigma_m\, S(mf_0)e^{i\varphi(mf_0)}e^{im\omega_0 t}. \qquad (2)$$

Third Case: Frequencies above f_c are not transmitted and the impulse response vanishes for $t \geq t_c$.—This is the combination of cases 1 and 2. It occurs approximately when discontinuities of the time and frequency functions are small.[1,3] In this case redundancy is eliminated if both the impulse response and the frequency-phase characteristics are replaced by discontinuous curves as explained for each case above. Equations (1) and (2) can be used, except that in Eq. (1) ω should be replaced by $m_1\omega_0$ and in Eq. (2) t should be replaced by $m_2 t_0$ where $\omega_0 = 2\pi f_0 = 2\pi/t_c$ and $t_0 = 1/2f_c$, and where m_1 and m_2 are integers such that $m_1\omega_0 \leqslant 2\pi f_c$ and $m_2 t_0 \leqslant t_c$.

Application in Practice.—The third case is the most important in practice since a considerable amount of redundancy can then be eliminated. If the impulse response is given, t_c can be fixed easily for the approximation required, and f_c can be estimated by the frequency of the vanishing oscillations of the tail of the response. If the frequency-phase characteristic is given, f_c can be fixed easily, while estimation of t_c requires some experience of similar cases on the use of properties given in our earlier papers.[1,2,4]

* This note gives a solution to the unsolved problem mentioned in "Radio Progress during 1951," Proc. Inst. Radio Engrs, 401 (April, 1951).

† Consultant to F. E. M. Company, Montreal. Formerly with Defence Research Board, Ottawa, Canada.

[1] M. Levy, Can. J. Phys. 147–158 (March, 1952).

[2] M. Levy, "A simple method for converting time response of a system in frequency-phase characteristics and inversely," Defence Research Board Report, April, 1951 (to be published in Can. J. Phys., August, 1952).

[3] M. Levy, Proc. Natl. Electronic Conference, 1951, p. 73 (abstract only).

[4] M. Levy, J. Brit. Inst. Radio Engr. VI, December, 1946.

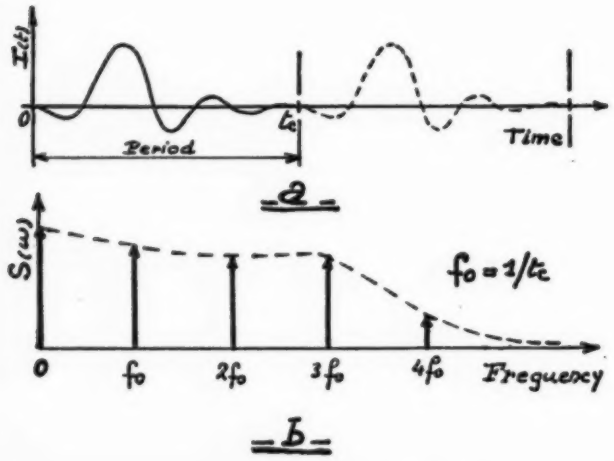


FIG. 2.

Random Noise in Dielectrics

HERBERT BAUSS AND R. F. BOYER
The Dow Chemical Company, Midland, Michigan
(Received April 14, 1952)

WE have recently found that the fluctuating electric current observed[1] when a direct potential is applied across thin films of certain polymers tends to disappear on cooling the sample below the second-order transition temperature. This is the temperature where microbrownian motion of chain segments ceases.[2] Hence we believe these new results afford strong evidence in favor of our original hypothesis that this "noise" arises from the activated diffusion of ions through the polymer, with the rate of diffusion being regulated by the natural jump frequency of the polymer segments.

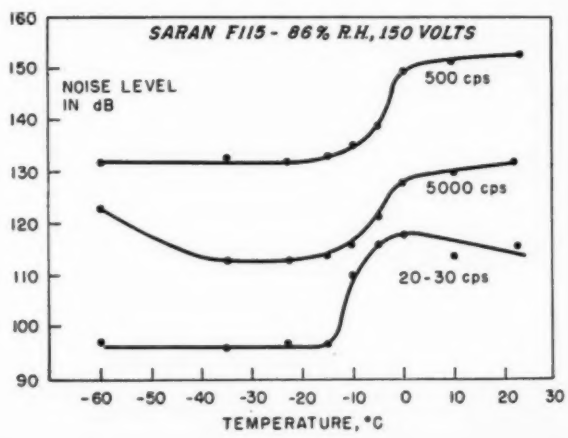


FIG. 1.

Figure 1 shows a typical plot obtained when the noise level in decibels is measured as a function of temperature at three different frequencies. The second-order transition temperature of this vinylidenechloride—vinylcyanide copolymer is −10°C. It is seen that the noise intensity changes an 100-fold at 20–30 cycles per second on going through this transition. The direct current conductivity (at 500 volts) increases only twofold, with rising temperature. At 500 cps, the transition has moved to a higher temperature, as is to be expected,[2] and has become more diffuse. At 5000 cps, the entire noise level has dropped, but still shows the transition. If noise is plotted as a function of frequency at 25°C, it shows a sharp maximum at 500 cps. The dielectric loss maximum for this same polymer at 25°C, is below 100 cps. Exactly similar results were obtained in the case of Nylon except that the break now occurs at 35° to 40°C, and at much higher frequencies, i.e., at 5 kc and higher. This is consistent with the dielectric measurements of Baker and Yager on Nylon.[3] The noise *vs* frequency curve for Nylon peaks at about 1100 cps, but the peak is relatively broad.

A full account of these results, including experimental techniques is being prepared.

[1] R. F. Boyer, J. Appl. Phys. **21**, 469 (1950).
[2] For a general discussion of second-order transition effects see R. F. Boyer and R. S. Spencer, in *Advances in Colloid Science*, edited by H. Mark and G. S. Whitby (Interscience Publishers, Inc., New York, 1946), Vol. 2, p. 1.
[3] W. O. Baker and W. A. Yager, J. Am. Chem. Soc. **64**, 2177 (1942)

JOURNAL OF APPLIED PHYSICS VOLUME 23, NUMBER 7 JULY, 1952

Announcements

Frank M. Folsom Scholarship Established at University of Notre Dame

An undergraduate scholarship for students majoring in pure science or engineering has been established by the Radio Corporation of America at the University of Notre Dame. The award provides an annual grant of $800 and will be known as the "Frank M. Folsom Scholarship" in honor of the President of RCA.

Candidates for the Frank M. Folsom Scholarship will be selected by academic officers of Notre Dame and recommended to the RCA Education Committee for consideration. The first recipient of the award will be announced during the Fall of 1952.

Research, Fellowship Grants Renewed by Shell Oil Company for 1952-1953

An increase of approximately 25 percent in grants for fellowships, stipends, and research expense allowances has been announced by the Shell organization as part of its renewal for the 1952–1953 academic year of its program in support of higher education. In recent years, industry has become increasingly aware of the value of extending aid for basic research, realizing that national defense and continued progress in industry are both vitally dependent on the advancements of science. The Shell program was established in recognition of the fact that although colleges and universities are well equipped to carry on such research, they are often handicapped by lack of the necessary funds.

Shell's graduate fellowship program is designed to help outstanding students obtain advanced scientific degrees. It provides awards to universities in the fields of chemistry, chemical engineering, geology, geophysics, mechanical engineering, petroleum production engineering, physics, and plant science. Students recommended by the colleges as Shell fellows receive a stipend of $1500 for the academic year, plus payment of tuition and fees, and the school receives $400 for related research expenses of the fellow. Recipients of these grants are under no obligation to Shell.

Both the fellowship and the basic research programs are administered by the Shell fellowship committee, made up of senior executives representing Shell Oil Company, Shell Chemical Corporation, Shell Development Company, and Shell Pipe Line Corporation.

Summer Courses, Conferences, Meetings

The American Oil Chemists Society Short Course, in cooperation with Rutgers University this summer, has as part of its program four industrial trips. These have been scheduled for Monday, July 7th, Tuesday July 8th, Thursday July 10th, and Friday July 11th, starting at 2:00 P.M. Each trip will be limited to about 50 people. Plants to be visited are: Lever Brothers, Edgewater, New Jersey, J. Howard Smith Company, Port Newark, New Jersey, Woburn Chemical Corporation, Kearny, New Jersey, Colgate-Palmolive-Peet Company, Jersey City, New Jersey. The balance of the details to be announced later by Dr. Foster Dee Snell, General Chairman, includes 4 speakers each morning for 5 days and dinner speakers 4 evenings of the week.

Further details as to the course are obtainable from the office of the Executive Secretary, 35 E. Wacker Drive, Chicago 1, Illinois.

Course in Modern Industrial Spectrography at Boston College

Boston College has announced a special two-weeks intensive course in Modern Industrial Spectrography at Chestnut Hill, Boston, Massachusetts, from July 21 to August 1. The course is particularly designed for chemists and physicists from industries in the process of installing spectrographic equipment. Information on the course can be obtained from Professor James J. Devlin, S. J., Physics Department, Boston College, Chestnut Hill 67, Boston, Massachusetts.

Course in Instrumental Analysis at M.I.T.

Two one-week specialized training programs in Instrumental Analysis will be offered from August 18 to 22 and from August 25 to 29 as part of the 1952 Summer Session at the Massachusetts Institute of Technology.

The courses are planned to enable chemists from industrial and governmental laboratories to study the applications of new instrumental techniques and methods in the field of applied analytical chemistry.

Since enrollment in these special courses will be limited by the availability of laboratory equipment, early application is advisable. Requests for application forms or for further information on these and other special summer activities at M.I.T. should be addressed to Dr. Ernest H. Huntress, Director of the Summer Session, Room 3–107, Massachusetts Institute of Technology, Cambridge 39, Massachusetts.

Summer Symposium at Oak Ridge

The Fourth Annual Oak Ridge Summer Symposium, scheduled for August 25–29, 1952, will be subtitled "The Role of Atomic Energy in Agricultural Research."

The symposium is being given by the Oak Ridge Institute of Nuclear Studies and the Oak Ridge National Laboratory under the sponsorship of the University of Tennessee-Atomic Energy Commission Agricultural Research Program.

Additional information may be obtained from the University Relations Division of the Institute, P. O. Box 117, Oak Ridge, Tennessee.

ONR Sponsored Conference on Magnetism

The Office of Naval Research is sponsoring a Conference on Magnetism to be held at the University of Maryland from September 2 to 6, 1952. Several foreign visitors, among them Professors Neel, Stoner, Bates, Sucksmith, and Gorter, are expected to attend and present papers. The conference will cover the following broad subjects: Technical Magnetization, Exchange Forces, Antiferromagnetism, Ferrites, Experimental Techniques, and Paramagnetism. An evening session devoted to a round table discussion of exchange phenomena and theories of the origin of ferromagnetism with the participation of Professors Neel, Slater, Stoner, Van Vleck, and Zener is also projected as a part of the program. The conference is being organized by the ONR magnetism panel whose Chairman is Professor R. Smoluchowski. The program committee consists of R. M. Bozorth, J. E. Goldman, G. T. Rado, and J. H. Van Vleck, Chairman. A local committee has been organized in Washington, D. C. to arrange for the meeting comprising E. Montroll, L. R. Maxwell, L. Marton, and G. T. Rado, Chairman. Inquiries concerning attendance and housing should be directed to Dr. Rado at the Naval Research Laboratory, Washington, D. C. It is expected that the foreign visitors will spend some time in this country visiting various centers of magnetic research. Communications concerning the scheduling of their visits to laboratories or universities should be addressed to Dr. V. Wilson, General Electric Research Laboratory, Schenectady, New York.

Annual Meeting of the BPA

The 22nd Annual Meeting of the Biological Photographic Association, Inc., will be held at the Hotel New Yorker, New York City, from September 10th through September 12th (Wednesday through Friday).

The Meeting of the BPA, an organization composed of medical, biological, clinical, scientific, and research photographers, will be conducted under the auspices of the New York Chapter of this national group.

Advance registration blanks for nonmembers may be obtained from the National Secretary of the Biological Photographic Association, Inc....Mr. Lloyd E. Varden, % Pavelle Color Incorporated, 533 West 57th Street, New York 19, New York.

National Electronics Conference. 8th annual conference. September 29, 30, and October 1, 1952, Sherman Hotel, Chicago, Illinois

The conference is sponsored by the American Institute of Electrical Engineers, Institute of Radio Engineers, Illinois Institute of Technology, Northwestern University, and the University of Illinois, with participation by Purdue University, University of Wisconsin, and the Society of Motion Picture & Television Engineers.

Recent Appointments

Dr. George A. Busch, formerly physicist at the Swiss Federal Institute of Technology, has been appointed visiting Professor of Physics at Carnegie Institute of Technology. An experimental and theoretical physicist, Dr. Busch will work mainly on Carnegie's program of research in low temperature phenomena.

Dr. L. E. Grinter, formerly vice president of Illinois Institute of Technology, has announced his appointment as Dean of the Graduate School and Director of Research at the University of Florida, Gainesville, Florida. Dean Grinter expects to take up his new duties at the beginning of the fall semester.

Dr. Kenneth L. Yudowitch, project chairman for the Operations Research Office, Johns Hopkins University, has been appointed supervisor of the physics of solids section at Armour Research Foundation of Illinois Institute of Technology. Dr. Yudowitch is known for research in x-ray diffraction, scattering, and instrumentation.

O. H. Olson, formerly research assistant for the Institute of Paper Chemistry, Appleton, Wisconsin, has been appointed a physicist in the physics department at Armour Research Foundation of Illinois Institute of Technology, and **William A. Casler,** assistant director of research, has been named manager of program development.

The Department of Defense Research and Development Board announces the following appointments: **Dr. J. B. Austin,** director of research and chief of United States Steel Company's Research Laboratory at Kearny, N. J., as member of the Committee on Chemical Warfare; **Ervin G. Bailey** and **Dr. Linton E. Grinter** as members of the Committee on Equipment and Supplies; **Dr. Hans A. Bethe** and **Harry A. Winne** as members of the Committee on Atomic Energy; **Earl G. Droessler** as executive director of the Committee on Geophysics and Geography; **Major General Kenneth D. Nichols,** Deputy Director of Guided Missiles within the Office of the Secretary of Defense, has been named Army service member of RDB; **Earl D. Johnson** has been designated as the civilian Army member of RDB; **Dr. Norman A. Shepard,** of American Cyanamid Company, has been appointed as Chairman of a Committee on Materials, and Dr. C. Guy Suits, vice president and director of research for General Electric, Schenectady, N. Y., as a civilian member of the Committee on Electronics.

Dr. Clifford E. Berry has been promoted to the office of Assistant Director of Research at Consolidated Engineering Corporation of Pasadena, California. Dr. Berry's particular distinction derives from his development of the Isatron, considered the heart of the mass spectrometer.

Bausch & Lomb Optical Company announces the appointment of **Heinz E. R. Gruner** as head of its Photogrammetric Section. A graduate of the University of Dresden, Germany, Gruner joined the firm's Scientific Bureau in 1948.

Dr. Raymond John Seeger, formerly Chief of the Aeroballistics Research Department of Ordnance Laboratory at White Oak, Maryland, has been appointed to the staff of the Division of Mathematical, Physical and Engineering Sciences of the National Science Foundation.

The National Science Foundation announces the following appointed members of the Divisional Committee for Mathematical, Physical, and Engineering Sciences: **Dr. A. Adrian Albert,** Professor of Mathematics, University of Chicago, Chicago, Illinois; **Professor Jesse W. Beams,** Chairman, School of Physics, University of Virginia, Charlottesville, Virginia; **Dr. William L. Everitt,** Dean, College of Engineering, University of Illinois, Urbana, Illinois; **Dr. Leo Goldberg,** Chairman, Department of Astronomy, University of Michigan, Ann Arbor, Michigan; **Professor Morrough P. O'Brien,** Chairman, Department of Engineering, University of California, Berkeley, California.

The Southwest Research Institute has appointed **Dr. Robert J. Anderson** as head of the recently established department of metallurgy; **Dr. C. A. Culver,** Senior Research Physicist, as Dean of Professional Development; **Dale Dorn,** vice president of the Forest Oil Corporation, has been elected chairman of the board of trustees for the Southwest Foundation for Research and Education; **Dr. Louis Koenig** has been named an associate director of Southwest Research Institute; also, **Norman C. Penfold.**